REPEAT SEQUENCES OF THE CA125 GENE AND THEIR USE FOR DIAGNOSTIC AND THERAPEUTIC INTERVENTIONS

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/284,175 filed April 17, 2001 and U.S. Provisional Application Serial No. 60/299,380 filed June 19, 2001, which are incorporated by reference in their entirety.

BACKGROUND OF THE INVENTION

The present invention relates generally to the cloning, identification, and expression of multiple repeat sequences of the CA125 gene *in vitro* and, more specifically, to the use of recombinant CA125 with epitope binding sites for diagnostic and therapeutic purposes.

CA125 is an antigenic determinant located on the surface of ovarian carcinoma cells with essentially no expression in normal adult ovarian tissue. Elevated in the sera of patients with ovarian adenocarcinoma, CA125 has played a critical role for more than 15 years in the management of these patients relative to their response to therapy and also as an indicator of recurrent disease.

It is well established that CA125 is not uniquely expressed in ovarian carcinoma, but is also found in both normal secretory tissues and other carcinomas (i.e., pancreas, liver, colon) [Hardardottir H *et al.*, Distribution of CA125 in embryonic tissue and adult derivatives of the fetal periderm, *Am J Obstet. Gynecol.* 163;6(1):1925-1931 (1990); Zurawski VR *et al.*, Tissue distribution and characteristics of the CA125 antigen, *Cancer Rev.* 11-12:102-108 (1988); and O'Brien TJ *et al.*, CA125 antigen in human amniotic fluid and fetal membranes, *Am J Obstet Gynecol.* 155:50-55, (1986); Nap M *et al.*, Immunohistochemical characterization of 22 monoclonal antibodies against the CA125 antigen: 2nd report from the ISOBM TD-1 workshop, Tumor Biology 17:325-332 (1996)]. Notwithstanding, CA125 correlates directly with the disease status of affected patients (i.e., progression, regression, and no change), and has become the "gold standard" for monitoring patients with ovarian carcinoma [Bast RC *et al.*, A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer, *N Engl J Med.* 309:883-887 (1983); and Bon GC *et al.*, Serum tumor marker

immunoassays in gynecologic oncology: Establishment of reference values, *Am J Obstet. Gynecol.* 174:107-114 (1996)]. CA125 is especially useful in post-menopausal patients where endometrial tissue has become atrophic and, as a result, is not a major source of normal circulating CA125.

During the mid 1980's, the inventor of the present invention and others developed M11, a monoclonal antibody to CA125. M11 binds to a dominant epitope on the repeat structure of the CA125 molecule [O'Brien TJ *et al.*, New monoclonal antibodies identify the glycoprotein carrying the CA125 epitope, *Am J Obstet Gynecol* 165:1857-64 (1991)]. More recently, the inventor and others developed a purification and stabilization scheme for CA125, which allows for the accumulation of highly purified high molecular weight CA125 [O'Brien TJ *et al.*, More than 15 years of CA125: What is known about the antigen, its structure and its function, *Int J Biological Markers* 13(4):188-195 (1998)].

Considerable progress has been made over the years to further characterize the CA125 molecule, its structure and its function. The CA125 molecule is a high molecular weight glycoprotein with a predominance of O-linked sugar side chains. The native molecule exists as a very large complex (~2-5 million daltons). The complex appears to be composed of an epitope containing CA125 molecule and binding proteins which carry no CA125 epitopes. The CA125 molecule is heterogenous in both size and charge, most likely due to continuous deglycosylation of the side chains during its life-span in bodily fluids. The core CA125 subunit is in excess of 200,000 daltons, and retains the capacity to bind both OC125 and M11 class antibodies. While the glycoprotein has been described biochemically and metabolically by the inventor of the present invention and others, no one has yet cloned the CA125 gene, which would provide the basis for understanding its structure and its physiologic role in both normal and malignant tissues.

Despite the advances in detection and quantitation of serum tumor markers like CA125, the majority of ovarian cancer patients are still diagnosed at an advanced stage of the disease--Stage III or IV. Further, the management of patients' responses to treatment and the detection of disease recurrence remain major problems. There, thus, remains a need to significantly improve and standardize current CA125 assay systems. Further, the development of an early indicator of risk of ovarian cancer will provide a useful tool for early diagnosis and improved prognosis.

SUMMARY OF THE INVENTION

The CA125 gene has been cloned and multiple repeat sequences as well as the carboxy terminus have been identified. CA125 requires a transcript of more than 35,000 bases and occupies approximately 150,000 bp on chromosome 19q 13.2. The CA125 molecule comprises three major domains: an extracellular amino terminal domain (Domain 1); a large multiple repeat domain (Domain 2); and a carboxy terminal domain (Domain 3) which includes a transmembrane anchor with a short cytoplasmic domain. The amino terminal domain is assembled by combining five genomic exons, four very short amino terminal sequences and one extraordinarily large exon. This domain is dominated by its capacity for O-glycosylation and its resultant richness in serine and threonine residues.

The extracellular repeat domain, which characterizes the CA125 molecule, also represents a major portion of the CA125 molecular structure. It is downstream from the amino terminal domain and presents itself in a much different manner to its extracellular matrix neighbors. These repeats are characterized by many features including a highly-conserved nature and a uniformity in exon structure. But most consistently, a cysteine enclosed sequence may form a cysteine loop. Domain 2 comprises 156 amino acid repeat units of the CA125 molecule. The repeat domain constitutes the largest proportion of the CA125 molecule. The repeat units also include the epitopes now well-described and classified for both the major class of CA125 antibodies of the OC125 group and the M11 group. More than 60 repeat units have been identified, sequenced, and contiguously placed in the CA125 domain structure. The repeat sequences demonstrated 70-85% homology to each other. The existence of the repeat sequences was confirmed by expression of the recombinant protein in *E. coli* where both OC125/M11 class antibodies were found to bind to sites on the CA125 repeat.

The CA125 molecule is anchored at its carboxy terminal through a transmembrane domain and a short cytoplasmic tail. The carboxy terminal also contains a proteolytic cleavage site approximately 50 amino acids upstream from the transmembrane domain, which allows for proteolytic cleavage and release of the CA125 molecule.

The identification and sequencing of multiple repeat domains of the CA125 antigen provides potentially new clinical and therapeutic applications for detecting, monitoring and treating patients with ovarian cancer and other carcinomas where CA125 is expressed. For

example, the ability to express repeat domains of CA125 with the appropriate epitopes would provide a much needed standard reagent for research and clinical applications. Current assays for CA125 utilize as standards either CA125 produced from cultured cell lines or from patient ascites fluid. Neither source is defined with regard to the quality or purity of the CA125 molecule. The present invention overcomes the disadvantages of current assays by providing multiple repeat domains of CA125 with epitope binding sites. At least one or more of any of the more than 60 repeats shown in Table 16 can be used as a "gold standard" for testing the presence of CA125. Furthermore, new and more specific assays may be developed utilizing recombinant products for antibody production.

Perhaps even more significantly, the multiple repeat domains of CA125 or other domains could also be used for the development of a potential vaccine for patients with ovarian cancer. In order to induce cellular and humoral immunity in humans to CA125, murine antibodies specific for CA125 were utilized in anticipation of patient production of anti-ideotypic antibodies, thus indirectly allowing the induction of an immune response to the CA125 molecule. With the availability of recombinant CA125, especially domains which encompass epitope binding sites for known murine antibodies, it will be feasible to more directly stimulate patients' immune systems to CA125 and, as a result, extend the life of ovarian carcinoma patients.

The recombinant CA125 of the present invention may also be used to develop therapeutic targets. Molecules like CA125, which are expressed on the surface of tumor cells, provide potential targets for immune stimulation, drug delivery, biological modifier delivery or any agent which can be specifically delivered to ultimately kill the tumor cells. Humanized or human antibodies to CA125 epitopes could be used to deliver all drug or toxic agents including radioactive agents to mediate direct killing of tumor cells. Natural ligands having a natural binding affinity for domains on the CA125 molecule could also be utilized to deliver therapeutic agents to tumor cells.

CA125 expression may further provide a survival or metastatic advantage to ovarian tumor cells. Antisense oligonucleotides derived from the CA125 repeat sequences could be used to down-regulate the expression of CA125. Further, antisense therapy could be used in association with a tumor cell delivery system of the type described above.

Recombinant domains of the CA125 molecule also have the potential to identify small molecules, which bind to individual domains of the CA125 molecule. These small molecules could also be used as delivery agents or as biological modifiers.

In one aspect of the present invention, a CA125 molecule is disclosed comprising: (a) an extracellular amino terminal domain, comprising 5 genomic exons, wherein exon 1 comprises amino acids #1-33 of SEQ ID NO: 299, exon 2 comprises amino acids #34-1593 of SEQ ID NO: 299, exon 3 comprises amino acids #1594-1605 of SEQ ID NO: 299, exon 4 comprises amino acids #1606-1617 of SEQ ID NO: 299, and exon 5 comprises amino acids #1618-1637 of SEQ ID NO: 299; (b) a multiple repeat domain, wherein each repeat unit comprises 5 genomic exons, wherein exon 1 comprises amino acids #1-42 in any of SEQ ID NOS: 164 through 194; exon 2 comprises amino acids #43-65 in any of SEQ ID NOS: 195 through 221; exon 3 comprises amino acids #66-123 in any of SEQ ID NOS: 222 through 249; exon 4 comprises amino acids #124-135 in any of SEQ ID NOS: 250 through 277; and exon 5 comprises amino acids #136-156 in any of SEQ ID NOS: 278 through 298; and (c) a carboxy terminal domain comprising a transmembrane anchor with a short cytoplasmic domain, and further comprising 9 genomic exons, wherein exon 1 comprises amino acids #1-11 of SEQ ID NO: 300; exon 2 comprises amino acids #12-33 of SEQ ID NO: 300; exon 3 comprises amino acids #34-82 of SEQ ID NO: 300; exon 4 comprises amino acids #83-133 of SEQ ID NO: 300; exon 5 comprises amino acids #134-156 of SEQ ID NO: 300; exon 6 comprises amino acids #157-212 of SEQ ID NO: 300; exon 7 comprises amino acids #213-225 of SEQ ID NO: 300; exon 8 comprises amino acids #226-253 of SEQ ID NO: 300; and exon 9 comprises amino acids #254-284 of SEQ ID NO: 300.

In another aspect of the present invention, the N-glycosylation sites of the amino terminal domain marked (x) in Figure 8B are encoded at positions #81, #271, #320, #624, #795, #834, #938, and #1,165 in SEQ ID NO: 299.

In another aspect of the present invention, the serine and threonine O-glycosylation pattern for the amino terminal domain is marked (o) in SEQ ID NO: 299 in Figure 8B.

In another aspect of the present invention, exon 2 in the repeat domain comprises at least 31 different copies; exon 2 comprises at least 27 different copies; exon 3 comprises at least 28 different copies; exon 4 comprises at least 28 different copies, and exon 5 comprises at least 21 different copies.

In another aspect of the present invention, the repeat domain comprises 156 amino acid repeat units which comprise epitope binding sites. The epitope binding sites are located in the C-enclosure at amino acids #59-79 (marked C-C) in SEQ ID NO: 150 in Figure 5.

In another aspect, the 156 amino acid repeat unit comprises O-glycosylation sites at positions #128, #129, #132, #133, #134, #135, #139, #145, #146, #148, #150, #151, and #156 in SEQ ID NO: 150 in Figure 5C. The 156 amino acid repeat unit further comprises N-glycosylation sites at positions #33 and #49 in SEQ ID NO: 150 in Figure 5C. The repeat unit also includes at least one conserved methionine (designated M) at position #24 in SEQ ID NO: 150 in Figure 5C.

In yet another aspect, the transmembrane domain of the carboxy terminal domain is located at positions #230-252 (underlined) in SEQ ID NO: 300 of Figure 9B. The cytoplasmic domain of the carboxy terminal domain comprises a highly basic sequence adjacent to the transmembrane at positions #256-260 in SEQ ID NO: 300 of Figure 9B, serine and threonine phosporylation sites at positions #254, #255, and #276 in SEQ ID NO: 300 in Figure 9B, and tyrosine phosphorylation sites at positions #264, #273, and #274 in SEQ ID NO: 300 of Figure 9B.

In another aspect of the present invention, an isolated nucleic acid of the CA125 gene is disclosed, which comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequences set forth in SEQ ID NOS: 49, 67, 81, 83-145, 147, 150, and 152; (b) a nucleotide sequence having at least 70% sequence identity to any one of the sequences in (a); (c) a degenerate variant of any one of (a) to (b); and (d) a fragment of any one of (a) to (c).

In another aspect of the present invention, an isolated nucleic acid of the CA125 gene, comprising a sequence that encodes a polypeptide with the amino acid sequence selected from the group consisting of: (a) the amino acid sequences set forth in SEQ ID NOS: 11-47, 50-80, 82, 146, 148, 149, 151, and 153-158; (b) an amino acid sequence having at least 50% sequence identity to any one of the sequences in (a); (c) a conservative variant of any one of (a) to (b); and (d) a fragment of any one of (a) to (c).

In yet another aspect, a vector comprising the nucleic acid of the CA125 gene is disclosed. The vector may be a cloning vector, a shuttle vector, or an expression vector. A cultured cell comprising the vector is also disclosed.

In yet another aspect, a method of expressing CA125 antigen in a cell is disclosed, comprising the steps of: (a) providing at least one nucleic acid comprising a nucleotide sequence selected from the group consisting of: (i) the nucleotide sequences set forth in SEQ ID NOS: 49, 67, 81, 83-145,

147, 150, and 152; (ii) a nucleotide sequence having at least 70% sequence identity to any one of the sequences in (i); (iii) a degenerate variant of any one of (i) to (ii); and (iv) a fragment of any one of (i) to (iii); (b) providing cells comprising an mRNA encoding the CA125 antigen; and (c) introducing the nucleic acid into the cells, wherein the CA125 antigen is expressed in the cells.

In yet another aspect, a purified polypeptide of the CA125 gene, comprising an amino acid sequence selected from the group consisting of: (a) the amino acid sequences set forth in SEQ ID NOS: 11-48, 50, 68-80, 82, 146, 148, 149, 150, 151, and 153-158; (b) an amino acid sequence having at least 50% sequence identity to any one of the sequences in (a); (c) a conservative variant of any one of (a) to (b); and (d) a fragment of any one of (a) to (c).

In another aspect, a purified antibody that selectively binds to an epitope in the receptor-binding domain of CA125 protein, wherein the epitope is within the amino acid sequence selected from the group consisting of: (a) the amino acid sequences set forth in SEQ ID NOS: 11-48, 50, 68-80, 146, 151, and 153-158; (b) an amino acid sequence having at least 50% sequence identity to any one of the sequences in (a); (c) a conservative variant of any one of (a) to (b); and (d) a fragment of any one of (a) to (c).

A diagnostic for detecting and monitoring the presence of CA125 antigen is also disclosed, which comprises recombinant CA125 comprising at least one repeat unit of the CA125 repeat domain including epitope binding sites selected from the group consisting of amino acid sequences set forth in SEQ ID NOS: 11-48, 50, 68-80, 82, 146, 150, 151, 153-161, and 162 (amino acids #1,643-11,438).

A therapeutic vaccine to treat mammals with elevated CA125 antigen levels or at risk of developing a disease or disease recurrence associated with elevated CA125 antigen levels is also disclosed. The vaccine comprises recombinant CA125 repeat domains including epitope binding sites, wherein the repeat domains are selected from the group of amino acid sequences consisting of SEQ ID NOS: 11-48, 50, 68-80, 82, 146, 148, 149, 150, 151, 153-161, and 162 (amino acids #1,643-11,438), and amino acids #175-284 of SEQ ID NO: 300. Mammals include animals and humans.

In another aspect of the present invention, an antisense oligonucleotide is disclosed that inhibits the expression of CA 125 encloded by: (a) the nucleotide sequences set forth in SEQ ID NOS: 49, 67, 81, 83-145, 147, 150, and 152; (b) a nucleotide sequence having at least 70% sequence identity to any one of the sequences in (a); (c) a degenerate variant of any one of (a) to (b); and (d) a fragment of any one of (a) to (c).

The preceeding and further aspects of the present invention will be apparent to those of ordinary skill in the art from the following description of the presently preferred embodiments of the invention, such description being merely illustrative of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates the cyanogen bromide digested products of CA125 on Western blot probed with M11 and OC125 antibodies. Table 1 shows the amino acid sequence derived from the amino terminal end of the 40 kDa cyanogen bromide peptide along with internal sequences obtained after protease digestion of the 40 kDa fragment (SEQ ID NOS: 1-4). SEQ ID NO: 1 is the amino terminal sequence derived of the 40 kDa peptide and SEQ ID NOS: 2, 3, and 4 reflect internal amino acid sequences derived from peptides after protease digestion of the 40 kDa fragment. Table 1 further provides a translation of the EST (BE005912) with homologous sequences (SEQ ID NOS: 5 and 6) either boxed or underlined. Protease cleavage sites are indicated by arrows.

Figure 2A illustrates PCR amplification of products generated from primers utilizing the EST sequence referred to in Figure 1, the amino acid sequence obtained from the 40 kDa fragment and EST sequence AA# 640762. Lane 1-2: normal; 3: serous ovarian carcinoma; 4: serous ovarian carcinoma; 5: mucinous ovarian carcinoma; 6: β-tubulin control. The anticipated size band 400 b is present in lane 3 and less abundantly in lane 4.

Figure 2B illustrates the RT-PCR that was performed to determine the presence or absence of CA125 transcripts in primary culture cells of ovarian tumors. This expression was compared to tubulin expression as an internal control. Lanes 1, 3, 5, 7, and 9 represent the primary ovarian tumor cell lines. Lanes 2, 4, 6, and 8 represent peripheral blood mononuclear cell lines derived from the corresponding patients in lanes 1, 3, 5, and 7. Lane 10 represents fibroblasts from the patient tumor in lane 9. Lanes 11 and 12 are CaOV3 and a primary tumor specimen, respectively.

Figure 3 illustrates repeat sequences determined by sequencing cloned cDNA from the 400 b band in Figure 2B. Placing of repeat sequences in a contiguous fashion was accomplished by PCR amplification and sequencing of overlap areas between two repeat sequences. A sample of the complete repeat sequences is shown in SEQ ID NOS: 158, 159, 160, and 161, which was obtained in this manner and placed next to each other based on overlap sequences. The complete list of repeat sequences that was obtained is shown in Table 21 (SEQ ID NO: 162).

Figure 4 illustrates three Western immunoblot patterns: Panel A = probed with M11, Panel B = probed with OC125 and Panel C = probed with antibody ISOBM 9.2. Each panel represents *E. coli* extracts as follows: lane 1 = *E. coli* extract from bacteria with the plasmid PQE-30 only. Lane 2 = *E. coli* extract from bacteria with the plasmid PQE-30 which includes the CA125 repeat unit. Lane 3 = *E. coli* extract from bacteria with the plasmid PQE-30 which includes the TADG-14 protease unrelated to CA125. Panel D shows a Coomassie blue stain of a PAGE gel of *E. coli* extract derived from either PQE-30 alone or from bacteria infected with PQE-30 - CA125 repeat (recombinant CA125 repeat).

Figure 5 represents Western blots of the CA125 repeat sequence that were generated to determine the position of the M11 epitope within the recombinant CA125 repeat. The expressed protein was bound to Ni-NTA agarose beads. The protein was left undigested or digested with Asp-N or Lys-C. The protein remaining bound to the beads was loaded into lanes 1, 2, or 3 corresponding to undigested, Asp-N digested and Lys-C digested, respectively. The supernatants from the digestions were loaded in lanes 4, 5, and 6 corresponding to undigested, Asp-N digested and Lys-C digested, respectively. The blots were probed with either anti-His tag antibody (A) or M11 antibody (B). Panel C shows a typical repeat sequence corresponding to SEQ ID NO: 150 with each exon defined by arrows. All proteolytic aspartic acid and lysine sites are marked with overhead arrow or dashes. In the lower panel, the O-glycosylation sites in exons 4 and 5 are marked with O, the N-glycosylation sites are marked with X plus the amino acid number in the repeat (#12, 33, and 49) the conserved methionine is designated with M plus the amino acid number (M#24), and the cysteine enclosure which is also present in all repeats and encompasses 19 amino acids between the cysteines is marked with C-C (amino acids #59-79). The epitopes for M11 and OC125 are located in the latter part of the C-enclosure or downstream from the C-enclosure.

Figure 6 illustrates a Northern blot analysis of RNA derived from either normal ovary (N) or ovarian carcinoma (T) probed with a P^{32} cDNA repeat sequence of CA125. Total RNA samples (10µg) were size separated by electrophoresis on a formaldehyde 1.2% agarose gel. After blotting to Hybond N, the lanes were probed with P^{32} radiolabelled 400 bp repeat (see Figure 2). Lane 1 represents RNA from normal ovarian tissue, and lane 2 represents RNA from serous ovarian tumor tissue.

Figure 7A is a schematic diagram of a typical repeat unit for CA125 showing the N-glycosylation sites at the amino end and the totally conserved methionine (M). Also shown is the proposed cysteine enclosed loop with antibody binding sites for OC125 and M11. Also noted are the highly O-glycosylated residues at the carboxy end of the repeat.

Figure 7B represents the genomic structure and exon configuration of a 156 amino acid repeat sequence of CA125 (SEQ ID NO: 163), which comprises a standard repeat unit.

Figure 7C lists the individual known sequences for each exon, which have been determined as follows: Exon 1 – SEQ ID NOS: 164-194; Exon 2 – SEQ ID NOS: 195-221; Exon 3 – SEQ ID NOS: 222-249; Exon 4 – SEQ ID NOS: 250-277; and Exon 5 – SEQ ID NOS: 278-298.

Figure 8A shows the genomic structure of the amino terminal end of the CA125 gene. It also indicates the amino composition of each exon in the extracellular domain.

Figure 8B illustrates the amino acid composition of the amino terminal domain (SEQ ID NO: 299) with each potential O-glycosylation site marked with a superscript (o) and N-glycosylation sites marked with a superscript (x). T-TALK sequences are underlined.

Figure 9A illustrates the genomic exon structure of the carboxy-terminal domain of the CA125 gene. It includes a diagram showing the extracellular portion, the potential cleavage site, the transmembrane domain and the cytoplasmic tail.

Figure 9B illustrates the amino acid composition of the carboxy terminal domain (SEQ ID NO: 300) including the exon boundaries, O-glycosylation sites (o), and N-glycosylation sites (x). The proposed transmembrane domain is underlined.

Figure 10 illustrates the proposed structure of the CA125 molecule based on the open reading frame sequence described herein. As shown, the molecule is dominated by a major repeat domain in the extracellular space along with a highly glycosylated amino terminal repeat. The molecule is anchored by a transmembrane domain and also includes a cytoplasmic tail with potential for phosphorylation.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the present invention, conventional molecular biology, microbiology, and recombinant DNA techniques may be used that will be apparent to those skilled in the relevant art. Such techniques are explained fully in the literature (see, e.g., Maniatis, Fritsch & Sambrook, "Molecular Cloning: A Laboratory Manual (1982); "DNA Cloning: A Practical

Approach," Volumes I and II (D. N. Glover ed. 1985); "Oligonucleotide Synthesis" (M. J. Gait ed. 1984); "Nucleic Acid Hybridization" (B. D. Hames & S. J. Higgins eds. (1985)); "Transcription and Translation" (B. D. Hames & S. J. Higgins eds. (1984)); "Animal Cell Culture" (R. I. Freshney, ed. (1986)); "Immobilized Cells And Enzymes" (IRL Press, (1986)); and B. Perbal, "A Practical Guide To Molecular Cloning" (1984)).

Therefore, if appearing herein, the following terms shall have the definitions set out below.

A "vector" is a replicon, such as plasmid, phage or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

A "DNA molecule" refers to the polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in either single stranded form, or a double-stranded helix. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, inter alia, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes.

As used herein, the term "gene" shall mean a region of DNA encoding a polypeptide chain.

"Messenger RNA" or "mRNA" shall mean an RNA molecule that encodes for one or more polypeptides.

"DNA polymerase" shall mean an enzyme which catalyzes the polymerization of deoxyribonucleotide triphosphates to make DNA chains using a DNA template.

"Reverse transcriptase" shall mean an enzyme which catalyzes the polymerization of deoxy- or ribonucleotide triphosphates to make DNA or RNA chains using an RNA or DNA template.

"Complementary DNA" or "cDNA" shall mean the DNA molecule synthesized by polymerization of deoxyribonucleotides by an enzyme with reverse transcriptase activity.

An "isolated nucleic acid" is a nucleic acid the structure of which is not identical to that of any naturally occurring nucleic acid or to that of any fragment of a naturally occurring genomic nucleic acid spanning more than three separate genes. The term therefore covers, for example, (a) a DNA which has the sequence of part of a naturally occurring genomic DNA molecule but is not flanked by both of the coding sequences that flank that part of the molecule in the genome of

the organism in which it naturally occurs; (b) a nucleic acid incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any naturally occurring vector or genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment; and (d) a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein.

"Oligonucleotide", as used herein in referring to the probes or primers of the present invention, is defined as a molecule comprised of two or more deoxy- or ribonucleotides, preferably more than ten. Its exact size will depend upon many factors which, in turn, depend upon the ultimate function and use of the oligonucleotide.

"DNA fragment" includes polynucleotides and/or oligonucleotides and refers to a plurality of joined nucleotide units formed from naturally-occurring bases and cyclofuranosyl groups joined by native phosphodiester bonds. This term effectively refers to naturally-occurring species or synthetic species formed from naturally-occurring subunits. "DNA fragment" also refers to purine and pyrimidine groups and moieties which function similarly but which have non naturally-occurring portions. Thus, DNA fragments may have altered sugar moieties or inter-sugar linkages. Exemplary among these are the phosphorothioate and other sulfur containing species. They may also contain altered base units or other modifications, provided that biological activity is retained. DNA fragments may also include species which include at least some modified base forms. Thus, purines and pyrimidines other than those normally found in nature may be so employed. Similarly, modifications on the cyclofuranose portions of the nucleotide subunits may also occur as long as biological function is not eliminated by such modifications.

"Primer" shall refer to an oligonucleotide, whether occurring naturally or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product, which is complementary to a nucleic acid strand, is induced, i.e., in the presence of nucleotides and an inducing agent such as a DNA polymerase and at a suitable temperature and pH. The primer may be either single-stranded or double-stranded and must be sufficiently long to prime the synthesis of the desired extension product in the presence of the inducing agent. The exact length of the primer will depend upon many factors, including temperature, the source of primer and the method used. For example, for

diagnostic applications, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 10-25 or more nucleotides, although it may contain fewer nucleotides.

The primers herein are selected to be "substantially" complementary to different strands of a particular target DNA sequence. This means that the primers must be sufficiently complementary to hybridize with their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence or hybridize therewith and thereby form the template for the synthesis of the extension product.

As used herein, the term "hybridization" refers generally to a technique wherein denatured RNA or DNA is combined with complementary nucleic acid sequence which is either free in solution or bound to a solid phase. As recognized by one skilled in the art, complete complementarity between the two nucleic acid sequences is not a pre-requisite for hybridization to occur. The technique is ubiquitous in molecular genetics and its use centers around the identification of particular DNA or RNA sequences within complex mixtures of nucleic acids.

As used herein, "restriction endonucleases" and "restriction enzymes" shall refer to bacterial enzymes which cut double-stranded DNA at or near a specific nucleotide sequence.

"Purified polypeptide"refers to any peptide generated from CA125 either by proteolytic cleavage or chemical cleavage.

"Degenerate variant" refers to any amino acid variation in the repeat sequence, which fulfills the homology exon structure and conserved sequences and is recognized by the M11, OC125 and ISOBM series of antibodies.

"Fragment" refers to any part of the CA125 molecule identified in a purification scheme.

"Conservative variant antibody" shall mean any antibody that fulfills the criteria of M11, OC125 or any of the ISOBM antibody series.

MATERIALS AND METHODS

A. Tissue collection, RNA Isolation and cDNA Synthesis

Both normal and ovarian tumor tissues were utilized for cDNA preparation. Tissues were routinely collected and stored at -80°C according to a tissue collection protocol.

Total RNA isolation was performed according to the manufacturer's instructions using the TriZol Reagent purchased from GibcoBRL (Catalog #15596-018). In some instances, mRNA was isolated using oligo dT affinity chromatography. The amount of RNA recovered was quantitated by UV spectrophotometry. First strand complementary DNA (cDNA) was synthesized using 5.0 μg of RNA and random hexamer primers according to the manufacturer's protocol utilizing a first strand synthesis kit obtained from Clontech (Catalog #K1402-1). The purity of the cDNA was evaluated by PCR using primers specific for the ß-tubulin gene. These primers span an intron such that the PCR products generated from pure cDNA can be distinguished from cDNA contaminated with genomic DNA.

B. Identification and Ordering of CA125 Repeat Units

It has been demonstrated that the 2-5 million dalton CA125 glycoprotein (with repeat domains) can be chemically segmented into glycopeptide fragments using cyanogen bromide. As shown in Figure 1, several of these fragments, in particular the 40 kDa and 60 kDa fragments, still bind to the to the two classical antibody groups defined by OC 125 and M11.

To convert CA125 into a consistent glycopeptide, the CA125 parent molecule was processed by cyanogen bromide digestion. This cleavage process resulted in two main fractions on commassie blue staining following polyacrylamide gel electrophoresis. An approximately 60 kDa band and a more dominant 40 kDa band were identified as shown in Figure 1. When a Western blot of these bands was probed with either OC125 or M11 antibodies (both of which define the CA125 molecule), these bands bound both antibodies. The 40 kDa band was significantly more prominent than the 60 kDa band. These data thus established the likelihood of these bands (most especially the 40 kDa band) as being an authentic cleavage peptide of the CA125 molecule, which retained the identifying characteristic of OC125 and M11 binding.

The 40 kDa and 60 kDa bands were excised from PVDF blots and submitted to amino terminal and internal peptide amino acid sequencing as described and practiced by Harvard Sequencing , (Harvard Microchemistry Facility and The Biological Laboratories, 16 Divinity

Avenue, Cambridge, Massachusetts 02138). Sequencing was successful only for the 40 kDa band where both amino terminal sequences and some internal sequences were obtained as shown in Table 1 at SEQ ID NOS: 1-4. The 40 kDa fragment of the CA125 protein was found to have homology to two translated EST sequences (GenBank Accession Nos. BE005912 and AA640762). Visual examination of these translated sequences revealed similar amino acid regions, indicating a possible repetitive domain. The nucleotide and amino acid sequences for EST Genbank Accession No. BE005912 (corresponding to SEQ ID NO: 5 and SEQ ID NO: 6, respectively) are illustrated in Table 1. Common sequences are boxed or underlined.

In an attempt to identify other individual members of this proposed repeat family, two oligonucleotide primers were synthesized based upon regions of homology in these EST sequences. Shown in Table 2A, the primer sequences correspond to SEQ ID NOS: 7 and 8 (sense primers) and SEQ ID NOS: 9 and 10 (antisense primers). Repeat sequences were amplified in accordance with the methods disclosed in the following references: Shigemasa K *et al.*, p21: A monitor of p53 dysfunction in ovarian neoplasia, *Int. J. Gynecol. Cancer* 7:296-303 (1997) and Shigemasa K *et al.,* p16 Overexpression: A potential early indicator of transformation in ovarian carcinoma, *J. Soc. Gynecol. Invest.* 4:95-102 (1997). Ovarian tumor cDNA obtained from a tumor cDNA bank was used.

Amplification was accomplished in a Thermal Cycler (Perkin-Elmer Cetus). The reaction mixture consisted of 1U Taq DNA Polymerase in storage buffer A (Promega), 1X Thermophilic DNA Polymerase 10X Mg free buffer (Promega), 300mM dNTPs, 2.5mM MgCl2, and 0.25mM each of the sense and antisense primers for the target gene. A 20 µl reaction included 1 µl of cDNA synthesized from 50ng of mRNA from serous tumor mRNA as the template. PCR reactions required an initial denaturation step at 94°C/1.5 min. followed by 35 cycles of 94°C/0.5 min., 48°C/0.5 min., 72°C/0.5 min. with a final extension at 72°C/7 min. Three bands were initially identified (»400 bp, »800 bp, and »1200 bp) and isolated. After size analysis by agarose gel electrophoresis, these bands as well as any other products of interest were then ligated into a T-vector plasmid (Promega) and transformed into competent DH5α strain of *E. coli* cells. After growth on selective media, individual colonies were cultured overnight at 37°C, and plasmid DNA was extracted using the QIAprep Spin Miniprep kit (Qiagen). Positive clones were identified by restriction digests using *Apa* I and *Sac* I. Inserts were sequenced using an ABI

automatic sequencer, Model 377, T7 primers, and a Big Dye Terminator Cycle Sequencing Kit (Applied Biosystems).

Obtained sequences were analyzed using the Pileup program of the Wisconsin Genetic's Computer Group (GCG). Repeat units were ordered using primers designed against two highly conserved regions within the nucleotide sequence of these identified repeat units. Shown in Table 2B, the sense and antisense primers (5'-GTCTCTATGTCAATGGTTTCACCC-3' / 5'-TAGCTGCTCTCTGTCCAGTCC-3' SEQ ID NOS: 301 and 302, respectively) faced away from one another within any one repeat creating an overlap sequence, thus enabling amplification across the junction of any two repeat units. PCR reactions, cloning, sequencing, and analysis were performed as described above.

C. Identification and Assembly of the CA125 Amino Terminal Domain

In search of open reading frames containing sequences in addition to CA125 repeat units, database searches were performed using the BLAST program available at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/). Using a repeat unit as the query sequence, cosmid AC008734 was identified as having multiple repeat sequences throughout the unordered (35) contiguous pieces of DNA, also known as contigs. One of these contigs, #32, was found to have exons 1 and 2 of a repeat region at its 3′ end. Contig#32 was also found to contain a large open reading frame (ORF) upstream of the repeat sequence. PCR was again used to verify the existence of this ORF and confirm its connection to the repeat sequence. The specific primers recognized the 3' end of this ORF (5'-CAGCAGAGACCAGCACGAGTACTC-3')(SEQ ID NO: 51) and sequence within the repeat (5'-TCCACTGCCATGGCTGAGCT-3')(SEQ ID NO: 52). The remainder of the amino-terminal domain was assembled from this contig in a similar manner. With each PCR confirmation, a new primer (see Table 10A) was designed against the assembled sequence and used in combination with a primer designed against another upstream potential ORF (Set 1: 5'-CCAGCACAGCTCTTCCCAGGAC-3' / 5'-GGAATGGCTGAGCTGACGTCTG-3'(SEQ ID NO: 53 and SEQ ID NO: 54); Set 2: 5'-CTTCCCAGGACAACCTCAAGG-3' / 5'-GCAGGATGAGTGAGCCACGTG-3'(SEQ ID NO: 55 and SEQ ID NO: 56); Set 3: 5'-GTCAGATCTGGTGACCTCACTG-3' / 5'-GAGGCACTGGAAAGCCCAGAG-3')(SEQ ID NO: 57 and SEQ ID NO: 58). Potential adjoining sequence (contig #7 containing EST AU133673) was also identified using contig #32 sequence as query sequence in database searches. Confirmation

primers were designed and used in a typical manner (5'-CTGATGGCATTATGGAACACATCAC-3' / 5'-CCCAGAACGAGAGACCAGTGAG-3')(SEQ ID NO: 59 and SEQ ID NO: 60).

In order to identify the 5' end of the CA125 sequence, 5' Rapid Amplification of cDNA Ends (FirstChoice™ RLM-RACE Kit, Ambion) was performed using tumor cDNA . The primary PCR reaction used a sense primer supplied by Ambion (5'-GCTGATGGCGATGAATGAACACTG-3') (SEQ ID NO: 61) and an anti-sense primer specific to confirmed contig #32 sequence (5'-CCCAGAACGAGAGACCAGTGAG-3')(SEQ ID NO: 62). The secondary PCR was then performed using nested primers, sense from Ambion (5'-CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG-3') (SEQ ID NO: 63) and the anti-sense was specific to confirmed contig #7 sequence (5'-CCTCTGTGTGCTGCTTCATTGGG-3')(SEQ ID NO: 64). The RACE PCR product (a band of approximately 300 bp) was cloned and sequenced as previously described.

D. Identification and Assembly of the CA125 Carboxy Terminal Domain

Database searches using confirmed repeat units as query also identified a cDNA sequence (GenBank AK024365) containing other repeat units, but also a potential carboxy terminal sequence. The contiguous nature of this sequence with assembled CA125 was confirmed using PCR (5'-GGACAAGGTCACCACACTCTAC-3' / 5'-GCAGATCCTCCAGGTCTAGGTGTG-3'), (SEQ ID NO: 303 and SEQ ID NO: 304, respectively) as well as contig and EST analysis.

E. Expression of 6xHis-tagged CA125 repeat in *E. coli*

The open reading frame of a CA125 repeat shown in Table 11 was amplified by PCR with the sense primer (5'-ACCGGATCCATGGGCCACACAGAGCCTGGCCC-3') (SEQ ID NO: 65) the antisense primer (5'-TGTAAGCTTAGGCAGGGAGGATGGAGTCC-3') (SEQ ID NO: 66) PCR was performed in a reaction mixture consisting of ovarian tumor cDNA derived from 50 ng of mRNA, 5 pmol each of sense and antisense primers for the CA125 repeat, 0.2 mmol of dNTPs, and 0.625 U of Taq polymerase in 1x buffer in a final volume of 25 ml. This mixture was subjected to 1 minute of denaturation at 95°C followed by 30 cycles of PCR consisting of the following: denaturation for 30 seconds at 95°C, 30 seconds of annealing at 62°C, and 1 minute of extension at 72°C with an additional 7 minutes of extension on the last cycle. The product was electrophoresed through a 2% agarose gel for separation. The PCR product was purified and digested with the restriction enzymes *Bam HI* and *Hind III*. This digested PCR product was then ligated into the expression vector pQE-30, which had also been digested with *Bam HI* and *Hind III*. This clone

would allow for expression of recombinant 6xHis-tagged CA125 repeat. Transformed *E. coli* (JM109) were grown to an OD600 of 1.5-2.0 at 37°C and then induced with IPTG (0.1 mM) for 4-6 hours at 25°C to produce recombinant protein. Whole *E. coli* lysate was electrophoresed through a 12% SDS polyacrylamide gel and Coomassie stained to detect highly expressed proteins.

F. Western Blot Analysis

Proteins were separated on a 12% SDS-PAGE gel and electroblotted at 100V for 40 minutes at 4°C to nitrocellulose membrane. Blots were blocked overnight in phosphate-buffered saline (PBS) pH 7.3 containing 5% non-fat milk. CA125 antibodies M11, OC125, or ISOBM 9.2 were incubated with the membrane at a dilution of 5µg/ml in 5% milk/PBS-T (PBS plus 0.1% TX-100) and incubated for 2 hours at room temperature. The blot was washed for 30 minutes with several changes of PBS and incubated with a 1:10,000 dilution of horseradish peroxidase (HRP) conjugated goat anti-mouse IgG antibody (Bio-Rad) for 1 hour at room temperature. Blots were washed for 30 minutes with several changes of PBS and incubated with a chemiluminescent substrate (ECL from Amersham Pharmacia Biotech) before a 10-second exposure to X-ray film for visualization.

Figure 4 illustrates three Western immunoblot patterns of the recombinant CA125 repeat purified from *E. coli* lysate (lane 2) compared to *E. coli* lysate with no recombinant protein (lane 1-negative control) and a recombinant protein TADG-14 which is unrelated to CA125 (lane 3). As shown, the M11 antibody, the OC125 antibody and the antibody ISOBM 9.2 (an OC125-like antibody) all recognized the CA125 recombinant repeat (lane 2), but did *not* recognize either the *E. coli* lysate (lane 1) or the unrelated TADG-14 recombinant (lane 3). These data confirm that the recombinant repeat encodes both independent epitopes for CA125, the OC125 epitope and the M11 epitope.

G. Northern Blot Analysis

Total RNA samples (approximately 10µg) were separated by electrophoresis through a 6.3% formaldehyde, 1.2% agarose gel in 0.02 M MOPS, 0.05 M sodium acetate (pH 7.0), and 0.001 M EDTA. The RNAs were then blotted to Hybond-N (Amersham) by capillary action in 20x SSPE and fixed to the membrane by baking for 2 hours at 80°C. A PCR product representing one 400 bp repeat of the CA125 molecule was radiolabelled using the Prime-a-Gene Labeling System available from Promega (cat. #U1100). The blot was probed and stripped

according to the ExpressHyb Hybridization Solution protocol available from Clontech (Catalog #8015-1).

RESULTS

In 1997, a system was described by a co-inventor of the present invention and others for purification of CA125 (primarily from patient ascites fluid), which when followed by cyanogen bromide digestion, resulted in peptide fragments of CA125 of 60 kDa and 40 kDa [O'Brien TJ *et al.*, More than 15 years of CA125: What is known about the antigen, its structure and its function, *Int J Biological Markers* 13(4)188-195 (1998)]. Both fragments were identifiable by commassie blue staining on polyacrylamide gels and by Western blot. Both fragments were shown to bind both OC125 and M11 antibodies, indicating both major classes of epitopes were preserved in the released peptides (Figure 1).

Protein sequencing of the 40 kDa band yielded both amino terminal sequences and some internal sequences generated by protease digestion (Table 1 – SEQ ID NOS: 1-4). Insufficient yields of the 60 kDa band resulted in unreliable sequence information. Unfortunately, efforts to amplify PCR products utilizing redundant primers designed to these sequences were not successful. In mid 2000, an EST (#BE005912) was entered into the GCG database, which contained homology to the 40 kDa band sequence as shown in Table 1 (SEQ ID NOS: 5 and 6). The translation of this EST indicated good homology to the amino terminal sequence of the 40 kDa repeat (e.g. PGSRKFKTTE) with only one amino acid difference (i.e. an asparagine is present instead of phenylalanine in the EST sequence). Also, some of the internal sequences are partially conserved (e.g. SEQ ID NO: 2 and to a lesser extent, SEQ ID NO: 3 and SEQ ID NO: 4). More importantly, all the internal sequences are preceded by a basic amino acid (Table 1, indicated by arrows) appropriate for proteolysis by the trypsin used to create the internal peptides from the 40 kDa cyanogen bromide repeat. Utilizing the combined sequences, those obtained by amino acid sequencing and those identified in the EST (#BE005912) and a second EST (#AA640762) identified in the database, sense primers were created as follows: 5'-GGA GAG GGT TCT GCA GGG TC-3' (SEQ ID NO: 7) representing amino acids ERVLQG and anti-sense primer, 5' GTG AAT GGT ATC AGG AGA GG-3' (SEQ ID NO: 9) representing PLLIPF. Using PCR, the presence of transcripts was confirmed representing these sequences in ovarian tumors and their absence in normal ovary and either very low levels or no detectable levels in a mucinous tumor (Figure 2A). The existence of transcripts was further

confirmed in cDNA derived from multiple primary ovarian carcinoma cell lines and the absence of transcripts in matched lymphocyte cultures from the same patient (Figure 2B).

After cloning and sequencing of the amplified 400 base pair PCR products, a series of sequences were identified, which had high homology to each other but which were clearly distinct repeat entities (Figure 3) (SEQ ID NOS: 158 through 161).

Examples of each category of repeats were sequenced, and the results are shown in Tables 3, 4, and 5. The sequences represent amplification and sequence data of PCR products obtained using oligonucleotide primers derived from an EST (Genbank Accession No. BE005912). Table 3 illustrates the amino acid sequence for a 400 bp repeat in the CA125 molecule, which is identified as SEQ ID NO: 11 through SEQ ID NO: 21. Table 4 illustrates the amino acid sequence for a 800 bp repeat in the CA125 molecule, which corresponds to SEQ ID NO: 22 through SEQ ID NO: 35. Table 5 illustrates the amino acid sequence for a 1200 bp repeat in the CA125 molecule, which is identified as SEQ ID NO: 36 through SEQ ID NO: 46. Assembly of these repeat sequences (which showed 75-80% homology to each other as determined by GCG Software (GCG = Genetics Computer Group) using the Pileup application) utilizing PCR amplification and sequencing of overlapping sequences allowed for the construction of a 9 repeat structure. The amino acid sequence for the 9 repeat is shown in Table 6 as SEQ ID NO: 47. The individual C-enclosures are highlighted in the table.

Using the assembled repeat sequence in Table 6 to search genebank databases, a cDNA sequence referred to as Genbank Accession No. AK024365 (entered on 9/29/00) was discovered. Table 7 shows the amino acid sequence for AK024365, which corresponds to SEQ ID NO: 48. AK024365 was found to overlap with two repeats of the assembled repeat sequence shown in Table 6. Individual C-enclosures are highlighted in Table 7.

The cDNA for AK024365 allowed alignment of four additional repeats as well as a downstream carboxy terminus sequence of the CA125 gene. Table 8 illustrates the complete DNA sequence of 13 repeats contiguous with the carboxy terminus of the CA125 molecule, which corresponds to SEQ ID NO: 49. Table 9 illustrates the complete amino acid sequence of the 13 repeats and the carboxy terminus of the CA125 molecule, which corresponds to SEQ ID NO: 50. The carboxy terminus domain was further confirmed by the existence of two EST's (Genbank Accession Nos. AW150602 and AI923224) in the genebank database, both of which

confirmed the stop-codon indicated (TGA) as well as the poly A signal sequence (AATAA) and the poly A tail (see Table 9). The presence of these repeats has been confirmed in serous ovarian tumors and their absence in normal ovarian tissue and mucinous tumors as expected (see Figure 2A). Also, the transcripts for these repeats have been shown to be present in tumor cell lines derived from ovarian tumors, but not in normal lymphocyte cell lines (Figure 2B). Moreover, Northern blot analysis of mRNA derived from normal or ovarian carcinoma and probed with a P^{32} labeled CA125 repeat sequence (as shown in Figure 6) confirmed the presence of an RNA transcript in excess of 20 kb in ovarian tumor extracts (see Figure 2B).

To date, 45 repeat sequences have been identified with high homology to each other. To order these repeat units, overlapping sequences were amplified using a sense primer (5' GTC TCT ATG TCA ATG GTT TCA CCC-3') (SEQ ID NO: 305) from an upstream repeat and an antisense primer from a downstream repeat sequence (antisense 5' TAG CTG CTC TCT GTC CAG TCC-3') (SEQ ID NO: 306). Attempts have been made to place these repeats in a contiguous fashion as shown in Figure 3. There is some potential redundancy. Further, there is evidence from overlapping sequences that some repeats exist in more than one location in the sequence giving a total of more than 60 repeats in the CA125 molecule (see Table 21 SEQ ID NO: 162).

Final confirmation of the relationship of the putative CA125 repeat domain to the known CA125 molecule was achieved by expressing a recombinant repeat domain in *E. coli*. In Figure 4, expression of a recombinant CA125 repeat domain is shown in lane 2 compared to the vector alone in lane 1, Panel D. A series of Western blots representing *E. coli* extracts of vector alone in lane 1; CA125 recombinant protein lane in 2 and recombinant TADG-14 (an unrelated recombinant protease), lane 3, were probed with the CA125 antibodies M11, Panel A; OC125, Panel B; and ISOBM 9.2, Panel C. In all cases, CA125 antibodies recognized only the recombinant CA125 antigen (lane 2 of each panel).

To further characterize the epitope location of the CA125 antibodies, recombinant CA125 repeat was digested with the endoprotease Lys-C and separately with the protease Asp-N. In both cases, epitope recognition was destroyed. As shown in Figure 5, the initial cleavage site for ASP-N is at amino acid #76 (indicated by arrow in Figure 5C). This sequence (amino acids # 1-76), a 17 kDa band, was detected with anti-histidine antibodies (Figure 5A,Lane 3) and found to have no capacity to bind CA125 antibodies (Figure 5B, Lane 3). The upper bands in Figures 5A and 5B represent the undigested remaining portion of the CA125 recombinant repeat. From these data, one

can reasonably conclude that epitopes are either located at the site of cleavage and are destroyed by Asp-N or are downstream from this site and also destroyed by cleavage. Likewise, cleavage with Lys-C would result in a peptide, which includes amino acids # 68-154 (Figure 5C) and again, no antibody binding was detected. In view of the foregoing, it seems likely that epitope binding resides in the cysteine loop region containing a possible disulfide bridge (amino acids # 59-79). Final confirmation of epitope sites are being examined by mutating individual amino acids.

To determine transcript size of the CA125 molecule, Northern blot analysis was performed on mRNA extracts from both normal and tumor tissues. In agreement with the notion that CA125 may be represented by an unusually large transcript due to its known mega dalton size in tumor sera, ascites fluid, and peritoneal fluid [Nustad K *et al.*, CA125 – epitopes and molecular size, *Int. J of Biolog.* Markers, 13(4)196-199 (1998)], a transcript was discovered which barely entered the gel from the holding well (Figure 6). CA125 mRNA was only present in the tumor RNA sample and while a precise designation of its true size remains difficult due to the lack of appropriate standards, its unusually large size would accommodate a protein core structure in excess of 11,000 amino acids.

Evidence demonstrates that the repeat domain of the CA125 molecule encompasses a minimum of 45 different 156 amino acid repeat units and possibly greater than 60 repeats, as individual repeats occur more than once in the sequence. This finding may well account for the extraordinary size of the observed transcript. The amino acid composition of the repeat units (Figure 7A, 7C, Table 21) indicates that the sequence is rich in serine, threonine, and proline typical of the high STP repeat regions of the mucin genes [Gum Jr., JR, Mucin genes and the proteins they encode: Structure, diversity and regulation, *Am J Respir. Cell Mol. Biol.* 7:557-564 (1992)]. Results suggest that the downstream end of the repeat is heavily glycosylated.

Also noteworthy is a totally conserved methionine at position 24 of the repeat (Figure 7A, 7C). It is this methionine which allowed cyanogen bromide digestion of the CA125 molecule, resulting in the 40 kDa glycopeptide that was identified with OC125 and M11 antibodies in Western blots of the CNBr digested peptides. These data predict that the epitopes for the CA125 antibodies are located in the repeat sequence. By production of a recombinant product representing the repeat sequence, results have confirmed this to be true. A potential disulfide bond is noted, which would encompass a C-enclosure comprising 19 amino acids enclosed by two cysteines at positions #59 and #79. The cysteines are totally conserved, which suggest a biological role for the resulting putative C-enclosure in each repeat. As mentioned above, it is likely that the OC125 and M11 epitopes are

located in the C-enclosure, indicating its relative availability for immune detection. This is probably due to the C-enclosure structure and the paucity of glycosylation in the immediate surrounding areas. Domain searches also suggest some homology in the repeat domain to an SEA domain commonly found in the mucin genes [Williams SJ *et al.*, MUC13, a novel human cell surface mucin expressed by epithelial and hemopoietic cells, *J of Biol. Chem* 276(21)18327-18336 (2001)] beginning at amino acid #1 and ending at #131 of each repeat. No biological function has been described for this domain.

Based on homology of the repeat sequences to chromosome 19q 13.2 (cosmid #AC008734) and confirmed by genomic amplification, it has been established that each repeat is comprised of 5 exons (covering approximately 1900 bases of genomic DNA): exon 1 comprises 42 amino acids (#1-42); exon 2 comprises 23 amino acids (#43-65); exon 3 comprises 58 amino acids (#66-123); exon 4 comprises 12 amino acids (#124-135); and exon 5 comprises 21amino acids (#136-156) (see Figure 7B). Homology pile-ups of individual exons have also been completed (see Figure 7C), which indicates that exon 1 has a minimum of 31different copies of the exon; exon 2 has 27 copies; exon 3 has 28 copies, exon 4 has 28 copies and exon 5 has 21 copies. If all exons were only found in a single configuration relative to each other, one could determine that a minimum number of repeats of 31 were present in the CA125 molecule. Using the exon 2 pile-up data as an example, it has been established as mentioned above that there are 27 individual exon 2 sequences. Using exon 2, which was sequenced fully in both the repeat units and the overlaps, results established that a minimum of 45 repeat units are present when exon 2 is combined with unique other exon combinations. However, based on overlap sequence information, 60+ repeat units are likely present in the CA125 molecule (Table 21). This larger number of repeat units can be accounted for by the presence of the same repeat unit occurring in more than one location.

Currently, the repetitive units of the repeat domain of the CA125 molecule constitute the majority of its extracellular molecular structure. These sequences have been presented in a tandem fashion based on overlap sequencing data. Some sequences may be incorrectly placed and some repeat units may not as yet be identified (Table 21). More recently, an additional repeat was identified in CA125 as shown in Tables 22 and 23 (SEQ. ID NOS: 307 and 308). The exact position has not yet been identified. Also, there is a potential that alternate splicing and/or mutation could account for some of the repeat variants that are listed. Studies are being conducted to compare both normal tissue derived CA125 repeats to individual tumor derived CA125 repeats to determine if such

variation is present. Currently, the known exon configurations would easily accommodate the greater than 60 repeat units as projected. It is, therefore, unlikely that alternate splicing is a major contributor to the repetitive sequences in CA125. It should also be noted that the genomic database for chromosome 19q 13.2 only includes about 10 repeat units, thus indicating a discrepancy between the data of the present invention (more than 60 repeats) and the genomic database. A recent evaluation of the methods used for selection and assembly for genomic sequence [Marshall E, DNA Sequencing: Genome teams adjust to shotgum marriage, *Science* 292:1982-1983 (2001)] reports that "more research is needed on repeat blocks of almost identical DNA sequence which are more common in the human genome. Existing assembly programs can't handle them well and often delete them." The CA125 repeat units located on chromosome 19 may well be victims of deletion in the genomic database, thus accounting for most CA125 repeat units absent from the current databases.

A. Sequence Confirmation and Assembly of the Amino Terminal Domain (Domain 1) of the CA125 Molecule

As previously mentioned, homology for repeat sequences was found in the chromosome 19 cosmid AC008734 of the GCG database. This cosmid at the time consisted of 35 unordered contigs. After searching the cosmid for repeat sequences, contig #32 was found to have exons 1 and 2 of a repeat unit at its 3' end. Contig #32 also had a large open reading frame upstream from the two repeat units, which suggested that this contig contained sequences consistent with the amino terminal end of the CA125 molecule. A sense primer was synthesized to the upstream non-repeat part of contig #32 coupled with a specific primer from within the repeat region (see Methods). PCR amplification of ovarian tumor cDNA confirmed the contiguous positioning of these two domains.

The PCR reaction yielded a band of approximately 980bp. The band was sequenced and found to connect the upstream open reading frame to the repeat region of CA125. From these data, more primer sets (see Methods) were synthesized and used in PCR reactions to piece together the entire open reading frame contained in contig #32. To find the 5' most end of the sequence, an EST (AU133673) was discovered, which linked contig #32 to contig #7 of the same cosmid. Specific primers were synthesized, (5'-CTGATGGCATTATGGAACACATCAC-3' (SEQ ID NO: 59) and 5'-CCCAGAACGAGAGACCAGTGAG-3' (SEQ ID NO: 60)), to the EST and contig #32. A PCR reaction was performed to confirm that part of the EST sequence was in fact contiguous with contig #32. Confirmation of this contiguous 5' prime sequencing strategy using overlapping sequences allowed the assembly of the 5' region (Domain 1) (Figure 8A). 5' RACE PCR was performed on

tumor cDNA to confirm the amino terminal sequence to CA125. The test confirmed the presence of contig #7 sequence at the amino terminal end of CA125.

The amino terminal domain comprises five genomic exons covering approximately 13,250 bp. Exon 1, a small exon, (amino acids #1-33) is derived from contig #7 (Figure 8A). The remaining exons are all derived from contig #32: Exon 2 (amino acids #34-1593), an extraordinarily large exon, Exon 3 (amino acids #1594-1605), Exon 4 (amino acids #1606-1617) and Exon 5 (amino acids #1618-1637) (see Figure 8A).

Potential N-glycosylation sites marked (x) are encoded at positions #81, #271, #320, #624, #795, #834, #938, and #1,165 (see Figure 8B). O-glycosylation sites are extraordinarily abundant and essentially cover the amino terminal domain (Figure 8B). As shown by the O-glycosylation pattern, Domain 1 is highly enriched in both threonine and serine (Figure 8B).

B. Sequence Confirmation and Assembly of the CA125 Carboxy Terminal End (Domain 3)

A search of Genbank using the repeat sequences described above uncovered a cDNA sequence referred to as Genbank accession number AK024365. This sequence was found to have 2 repeat sequences, which overlapped 2 known repeat sequences of a series of 6 repeats. As a result, the cDNA allowed the alignment of all six carboxy terminal repeats along with a unique carboxy terminal sequence. The carboxy terminus was further confirmed by the existence of two other ESTs (Genbank accession numbers AW150602 and A1923224), both of which confirmed a stop codon as well as a poly-A signal sequence and a poly-A tail (see GCG database #AF414442). The sequence of the carboxy terminal domain was confirmed using primers designed to sequence just downstream of the repeat domain (sense primer 5' GGA CAA GGT CAC CAC ACT CTA C-3') (SEQ ID NO: 303) and an antisense primer (5'-GCA GAT CCT CCA GGT CTA GGT GTG-3') (SEQ ID NO: 304) designed to carboxy terminus (Figure 9A).

The carboxy terminal domain covers more than 14,000 genomic bp. By ligation, this domain comprises nine exons as shown in Figure 9A. The carboxy-terminus is defined by a 284 amino acid sequence downstream from the repeat domains (see Figure 9B). Both N-glycosylation sites marked (x) (#31, #64, #103, #140, #194, #200) and a small number of O-glycosylation sites marked (o) are predicted for the carboxy end of the molecule (Figures 9A, 9B). Of special note is a putative transmembrane domain at positions #230-#252 followed by a cytoplasmic domain, which is characterized by a highly basic sequence adjacent to the membrane (#256-#260) as well as several

potential S/T phosphorylation sites (#254, #255, #276) and tyrosine phosphorylation sites (at # 264, #273, #274) (Figures 9A, 9B).

Assembly of the CA125 molecule as validated by PCR amplification of overlap sequence provides a picture of the whole molecule (see Figure 10 and Table 21). The complete nucleotide sequence is available in Genebank, Accession #AF414442 and the amino acid sequence as currently aligned is shown in Table 21.

DISCUSSION

The CA125 molecule comprises three major domains; an extracellular amino terminal domain (Domain 1), a large multiple repeat domain (Domain 2) and a carboxy terminal domain (Domain 3), which includes a transmembrane anchor with a short cytoplasmic domain (Figure 10). The amino terminal domain is assembled by combining five genomic exons, four very short amino terminal sequences and one extraordinarily large exon, which often typifies mucin extracellular glycosylated domains [Desseyn JL *et al.*, Human mucin gene MUC5B, the 10.7-kb large central exon encodes various alternate subdomains resulting in a super-repeat. Structural evidence for a 11p15.5 gene family, *J. Biol. Chem.* 272(6):3168-3178 (1997)]. This domain is dominated by its capacity for O-glycosylation and its resultant richness in serine and threonine residues. Overall, the potential for O-glycosylation essentially covers this domain and, as such, may allow the carbohydrate superstructure to influence ECM interaction at this end of the CA125 molecule (Figure 8). There is one short area (amino acids # 74-120) where little or no glycosylation is predicted, which could allow for protein-protein interaction in the extracellular matrix.

Efforts to purify CA125 over the years were obviously complicated by the presence of this amino terminal domain, which is unlikely to have any epitope sites recognized by the OC125 or M11 class antibodies. As the CA125 molecule is degraded *in vivo*, it is likely that this highly glycosylated amino terminal end will be found associated with varying numbers of repeat units. This could very well account for both the charge and size heterogeneity of the CA125 molecule so often identified from serum and ascites fluid. Also of note are two T-TALK sequences at amino acids # 45-58 (underlined in Figure 8B), which are unique to the CA125 molecule.

The extracellular repeat domain, which characterizes the CA125 molecule, also represents a major portion of the molecular structure. It is downstream from the amino terminal domain and presents itself in a much different manner to its extracellular matrix neighbors. These repeats are characterized by many features including a highly-conserved nature (Figure 3) and a uniformity in

exon structure (Figure 7). But most consistently, a cysteine enclosed sequence may form a cysteine loop (Table 21). This structure may provide extraordinary potential for interaction with neighboring matrix molecules. Domain 2 encompasses the 156 amino acid repeat units of the CA125 molecule. The repeat domain constitutes the largest proportion of the CA125 molecule (Table 21 and Figure 10). Because it has been known for more than 15 years that antibodies bind in a multivalent fashion to CA125, it has been predicted that the CA125 molecule would include multiple repeat domains capable of binding the OC125 and M11 class of sentinel antibodies which define this molecule [O'Brien *et al.*, New monoclonal antibodies identify the glycoprotein carrying the CA125 epitope, *Am J Obstet Gynecol.* 165:1857-1964 (1991); Nustad K *et al.*, Specificity and affinity of 26 monoclonal antibodies against the CA125 antigen: First report from the ISOBM TD-1 workshop, *Tumor Biology* 17:196-219 (1996); and Bast RC *et al.*, A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer, *N. Engl. J. Med.* 309:883-887 (1983)]. In the present invention, more than 60 repeat units have been identified, which are in tandem array in the extracellular portion of the CA125 molecule. Individual repeat units have been confirmed by sequencing and further identified by PCR amplification of the overlapping repeat sequences. Results confirm the contiguous placement of most repeats relative to its neighbor (Table 21).

Initial evidence suggests that this area is a potential site for antibody binding and also for ligand binding. The highly conserved methionine and several highly conserved sequences within the repeat domain also suggests a functional capacity for these repeat units. The extensive glycosylation of exons 4 & 5 of the repeat unit and the N-glycosylation potential in exon 1 and the 5' end of exon 2 might further point to a functional capacity for the latter part of exon 2 and exon 3 which includes the C-enclosure (see Figure 7). It should be apparent that the C-enclosure might be a prime target for protease activity and such cleavage may well explain the difficulty experienced by many investigators in obtaining an undigested CA125 parent molecule. Such activity might explain the diffuse pattern of antibody binding and the loss of antibody binding for molecules of less than 200,000 kDa. Proteolysis would destroy the epitopes and, therefore, only multiple repeats could be identified by blotting with CA125 antibodies. The repeat unit organization also suggests the potential for a multivalent interaction with extracellular entities.

The carboxy terminal domain of the CA125 molecule comprises an extracellular domain, which does not have any homology to other known domains. It encodes a typical transmembrane domain and a short cytoplasmic tail. It also contains a proteolytic cleavage site approximately 50

amino acids upstream from the transmembrane domain. This would allow for proteolytic cleavage and release of the CA125 molecule (Figure 9). As indicated by Fendrick, *et al.* [CA125 phosphorylation is associated with its secretion from the WISH human amnion cell line, *Tumor Biology* 18:278-289 (1997)], release of the CA125 molecule is preceded by phosphorylation and sustained by inhibitors of phosphatases, especially inhibition of phosphatase 2B. The cytoplasmic tail which contains S/T phosphorylation sites next to the transmembrane domain and tyrosine phosphorylation sites downstream from there could accommodate such phosphorylation. A very distinguishable positively charged sequence is present upstream from the tyrosine, suggesting a signal transduction system involving negatively charged phosphate groups and positively charged lysine and arginine groups.

These features of the CA125 molecule suggest a signal transduction pathway involvement in the biological function of CA125 [Fendrick JL *et al.*, CA125 phosphorylation is associated with its secretion from the WISH human amnion cell line, *Tumor Biology* 18:278-289 (1997); and Konish I *et al.*, Epidermal growth factor enhances secretion of the ovarian tumor-associated cancer antigen CA125 from the human amnion WISH cell line, *J Soc. Gynecol. Invest.* 1:89-96 (1994)]. It also reinforces the prediction of phosphorylation prior to CA125 release from the membrane surface as previously proposed [Fendrick JL *et al.*, CA125 phosphorylation is associated with its secretion from the WISH human amnion cell line, *Tumor Biology* 18:278-289 (1997); and Konish I *et al.*, Epidermal growth factor enhances secretion of the ovarian tumor-associated cancer antigen CA125 from the human amnion WISH cell line, *J Soc. Gynecol. Invest.* 1:89-96 (1994)]. Furthermore, a putative proteolytic cleavage site on the extra-cellular side of the transmembrane domain is present at position #176-181.

How well does the CA125 structure described in the present invention compare to the previously known CA125 structure? O'Brien *et al.* reported that a number of questions needed to be addressed: 1) the multivalent nature of the molecule; 2) the heterogeneity of CA125; 3) the carbohydrate composition; 4) the secretory or membrane bound nature of the CA125 molecule; 5) the function of the CA125 molecule; and 6) the elusive CA125 gene [More than 15 years of CA125: What is known about the antigen, its structure and its function, *Int J Biological Markers* 13(4)188-195 (1998)]. Several of these questions have been addressed in the present invention including, of course, the gene and its protein core product. Perhaps, most interestingly is the question of whether an individual large transcript accounted for the whole CA125 molecule, or a number of smaller

transcripts which represented subunits that specifically associated to produce the CA125 molecule. From the results produced by way of the present invention, it is now apparent that the transcript of CA125 is large - similar to some of the mucin gene transcripts e.g. MUC 5B [see Verma M *et al.*, Mucin genes: Structure, expression and regulation, *Glycoconjugate J.* 11:172-179 (1994); and Gendler SJ *et al.*, Epithelial mucin genes, *Annu. Rev. Physiol.* 57:607-634 (1995)]. The protein core extracellular domains all have a high capacity for O-glycosylation and, therefore, probably accounts for the heterogeneity of charge and size encountered in the isolation of CA125. The data also confirm the O-glycosylation inhibition data, indicating CA125 to be rich in O-glycosylation [Lloyd KO *et al.*, Synthesis and secretion of the ovarian cancer antigen CA125 by the human cancer cell line NIH: OVCAR-3, *Tumor Biology* 22, 77-82 (2001); Lloyd KO *et al.*, Isolation and characterization of ovarian cancer antigen CA125 using a new monoclonal antibody (VK-8): Identification as a mucin-type molecule, *Int. J. Cancer*, 71:842-850 (1997); and Fendrick JL *et al.*, Characterization of CA125 synthesized by the human epithelial amnion WISH cell line, *Tumor Biology* 14:310-318 (1993)].

The repeat domain which includes more than 60 repeat units accounts for the multivalent nature of the epitopes present, as each repeat unit likely contains epitope binding sites for both OC125-like antibodies and M11-like antibodies. The presence of a transmembrane domain and cleavage site confirms the membrane association of CA125, and reinforces the data which indicates a dependence of CA125 release on proteolysis. Also, the release of CA125 from the cell surface may well depend on cytoplasmic phosphorylation and be the result of EGF signaling [Nustad K *et al.*, Specificity and affinity of 26 monoclonal antibodies against the CA125 antigen: First report from the ISOBM TD-1 workshop, *Tumor Biology* 17:196-219 (1996)]. As for the question of inherent capacity of CA125 for proteolytic activity, this does not appear to be the case. However, it is likely that the associated proteins isolated along with CA125 (e.g. the 50 kDa protein which has no antibody binding ability) may have proteolytic activity. In any case, proteolysis of an extracellular cleavage site is the most likely mechanism of CA125 release. Such cleavage would be responsive to cytoplasmic signaling and mediated by an associated extracellular protease activity.

In summary, the large number of tandem repeats of the CA125 molecule, which dominate its molecular structure and contain the likely epitope binding sites of the CA125 molecule, was unexpected. Also, one cannot as yet account for the proteolytic activity, which has plagued the isolation and characterization of this molecule for many years. While no protease domain per se is constituitively part of the CA125 molecule, there is a high likelihood of a direct association by an

extracellular protease with the ligand binding domains of the CA125 molecule. Finally, what is the role of the dominant repeat domain of this extracellular structure? Based on the expression data of CA125 on epithelial surfaces and in glandular ducts, it is reasonable to conclude that the unique structure of these repeat units with their cysteine loops plays a role both as glandular anti-invasive molecules (bacterial entrapment) and/or a role in anti-adhesion (maintaining patency) between epithelial surfaces and in ductal linings.

Recently, Yin and Lloyd described the partial cloning of the CA125 antigen using a completely different approach to that described in the present invention [Yin TWT *et al.*, Molecular cloning of the CA125 ovarian cancer antigen. Identification as a new mucin (MUC16), *J Biol. Chem.* 276:27371-27375 (2001)]. Utilizing a polyclonal antibody to CA125 to screen an expression library of the ovarian tumor cell line OVCAR-3, these researchers identified a 5965 bp clone containing a stop codon and a poly A tail, which included nine partially conserved tandem repeats followed by a potential transmembrane region with a cytoplasmic tail. The 5965 bp sequence is almost completely homologous to the carboxy terminus region shown in Table 21. Although differing in a few bases, the sequences are homologous. As mentioned above, the cytoplasmic tail has the potential for phosphorylation and a transmembrane domain would anchor this part of the CA125 molecule to the surface of the epithelial or tumor cell. In the extracellular matrix, a relatively short transition domain connects the transmembrane anchor to a series of tandem repeats - in the case of Yin and Lloyd, nine.

By contrast, the major extracellular part of the molecule of the present invention as shown is upstream from the sequence described by Yin and includes a large series of tandem repeats. These results, of course, provide a different picture of the CA125 molecule, which suggest that CA125 is dominated by the series of extracellular repeats. Also included is a major amino terminal domain (~1638 amino acids) for the CA125 molecule, which it is believed accounts for a great deal of the O-glycosylation known to be an important structural component of CA125.

In conclusion, a CA125 molecule is disclosed which requires a transcript of more than 35,000 bases and occupies approximately 150,000 bp on chromosome 19q 13.2. It is dominated by a large series of extracellular repeat units (156 amino acids), which offer the potential for molecular interactions especially through a highly conserved unique cysteine loop. The repeat units also include the epitopes now well-described and classified for both the major class of CA125 antibodies (i.e., the OC125 and the M11 groups). The CA125 molecule is anchored at its carboxy terminal through a transmembrane domain and a short cytoplasmic tail. CA125 also contains a highly

glycosylated amino terminal domain, which includes a large extracellular exon typical of some mucins. Given the massive repeat domain presence of both epithelial surfaces and ovarian tumor cell surfaces, it might be anticipated that CA125 may play a major role in determining the extracellular environment surrounding epithelial and tumor cells.

Advantages and Uses of the CA125 Recombinant Products

1) Current assays to CA125 utilize as standards either CA125 produced from cultured cell lines or from patient ascites fluid. Neither source is defined with regard to the quality or purity of the CA125 molecule. Therefore arbitrary units are used to describe patient levels of CA125. Because cut-off values are important in the treatment of patients with elevated CA125 and because many different assay systems are used clinically to measure CA125, it is relevant and indeed necessary to define a standard for all CA125 assays. Recombinant CA125 containing epitope binding sites could fulfill this need for standardization. Furthermore, new and more specific assays may be developed utilizing recombinant products for antibody production.

2) Vaccines: Adequate data now exists [see Wagner U *et al.*, Immunological consolidation of ovarian carcinoma recurrences with monoclonal anti-idiotype antibody ACA125: Immune responses and survival in palliative treatment, *Clin. Cancer Res.* 7:1112-1115 (2001)], which suggest and support the idea that CA125 could be used as a therapeutic vaccine to treat patients with ovarian carcinoma. Heretofore, in order to induce cellular and humoral immunity in humans to CA125, murine antibodies specific for CA125 were utilized in anticipation of patient production of anti-ideotypic antibodies, thus indirectly allowing the induction of an immune response to the CA125 molecule. With the availability of recombinant CA125, especially domains which encompass epitope binding sites for known murine antibodies and domains directly anchoring CA125 on the tumor cell, it will be feasible to more directly stimulate patients' immune systems to CA125 and as a result, extend the life of ovarian carcinoma patients as demonstrated by Wagner et al.

Several approaches can be utilized to achieve such a therapeutic response in the immune system by: 1) directly immunizing the patient with recombinant antigen containing the CA125 epitopes or other domains; 2) harvesting dendritic cells from the patient; 3) expanding these cells in *in vitro* culture; 4) activating the dendritic cells with the recombinant CA125 epitope domain or other domains or with peptides derived from these domains [see Santin AD *et al.*, Induction of

ovarian tumor-specific CD8+ cytotoxic T lymphocytes by acid-eluted peptide-pulsed autologous dendritic cells, *Obstetrics & Gynecology* 96(3):422-430 (2000)]; and then 5) returning these immune stem cells to the patient to achieve an immune response to CA125. This procedure can also be accomplished using specific peptides which are compatible with histocompatibility antigens of the patient. Such peptides compatible with the HLA-A2 binding motifs common in the population are indicated in Figure 12.

3) Therapeutic Targets: Molecules, which are expressed on the surface of tumor cells as CA125 is, offer potential targets for immune stimulation, drug delivery, biological modifier delivery or any agent which can be specifically delivered to ultimately kill the tumor cells. CA125 offers such potential as a target: 1) Antibodies to CA125 epitopes or newly described potential epitopes: Most especially humanized or human antibodies to CA125 which could directly activate the patients' immune system to attack and kill tumor cells. Antibodies could be used to deliver all drug or toxic agents including radioactive agents to mediate direct killing of tumor cells. 2) Natural ligands: Under normal circumstances, molecules are bound to the CA125 molecule e.g. a 50 k dalton protein which does not contain CA125 epitopes co-purifies with CA125. Such a molecule, which might have a natural binding affinity for domains on the CA125 molecule, could also be utilized to deliver therapeutic agents to tumor cells.

4) Anti-sense therapy: CA125 expression may provide a survival or metastatic advantage to ovarian tumor cells as such antisense oligonucleotide derived from the CA125 sequence could be used to down-regulate the expression of CA125. Antisense therapy could be used in association with a tumor cell delivery system such as described above.

5) Small Molecules: Recombinant domains of CA125 also offer the potential to identify small molecules which bind to individual domains of the molecule. Small molecules either from combinatorial chemical libraries or small peptides can also be used as delivery agents or as biological modifiers.

All references referred to herein are hereby incorporated by reference in their entirety.

It should be understood that various changes and modifications to the presently preferred embodiments described herein will be apparent to those skilled in the art. Such changes and modifications can be made without departing from the spirit and scope of the present invention and without diminishing its attendant advantages.

TABLE 1

Comparison of the Amino Acid Terminal Sequences and Several Internal Sequences for the 40kD Band for CA125 glycoprotein (SEQ ID NO: 1 through SEQ ID NO: 4) to the Nucleotide and Amino Acid Sequences for EST Genbank Accession No. AA640762 (SEQ ID NO: 5 and SEQ ID NO: 6, respectively)

40kDa Nterm – QHPGSRKFKTTEG **(SEQ ID NO: 1)**

Peak 68 – FLTVERVLQGL **(SEQ ID NO: 2)**

Peak 65 – DTYVGPLY **(SEQ ID NO: 3)**

Peak 30 – DGAANGVD **(SEQ ID NO: 4)**

(SEQ ID NO: 5 and SEQ ID NO: 6)

```
                                    ↓
1   CGTCGACCTGGCTCTAGAAAGTTTAACACCACGGAGAGAGTCCTTCAGGGTCTGCTCAGG
    R  R  P  G  S  R  K  F  N  T  T  E  R  V  L  Q  G  L  L  R
                                    ↓
61  CCTGTGTTCAAGAACACCAGTGTTGGCCCTCTGTACTCTGGCTGCAGACTGACCTTGCTC
    P  V  F  K  N  T  S  V  G  P  L  Y  S  G  C  R  L  T  L  L
                                    ↓
121 AGGCCCAAGAAGGATGGGGCAGCCACCAAAGTGGATGCCATCTGCACCTACCGCCCTGAT
    R  P  K  K  D  G  A  A  T  K  V  D  A  I  C  T  Y  R  P  D

181 CCCAAAAGCCCTGGACTGGACAGAGAGCAGCTATACTGGGAGCTGAGCCAGGGTGATGCA
    P  K  S  P  G  L  D  R  E  Q  L  Y  W  E  L  S  Q  G  D  A
```

TABLE 2A

Nucleotide and Amino Acid Sequences for Sense Primer 5' 3' (SEQ ID NO: 7 and SEQ ID NO: 8 respectively) and Antisense Primer 5' 3' (SEQ ID NO: 9 and SEQ ID NO: 10 respectively) based upon Regions of Homology for EST Genbank Accession Nos. BE005912 and AA640762)

GGA GAG GGT TCT GCA GGG TC	**(SEQ ID NO: 7)**
E R V L Q G	**(SEQ ID NO: 8)**
GTG AAT GGT ATC AGG AGA GG	**(SEQ ID NO: 9)**
P L L I P F	**(SEQ ID NO: 10)**

TABLE 2B

Sense and Anti-Sense Primers Used for Ordering Repeat Units (SEQ ID NO: 301 and SEQ ID NO: 302, respectively)

5'-GTCTCTATGTCAATGGTTTCACCC-3'	**(SEQ ID NO: 301)**
5'-TAGCTGCTCTCTGTCCAGTCC-3'	**(SEQ ID NO: 302)**

TABLE 3

Amino Acid Sequence for a 400 bp Repeat in the CA125 Molecule (SEQ ID NO: 11 thru SEQ ID NO: 21)

	1			50	
12	ERVLQGLLRS	LFKSTSVGPL	YSGCRLTLLR	PEKDGTATGV DAICTHHPDP	(SEQ ID NO: 11)
34	ERVLQGLLMP	LFKNTSVSSL	YSGCRLTLLR	PEKDGAATRA DAVCTHRPDP	(SEQ ID NO: 12)
32	ERVLQGLLGP	IFKNTSVGPL	YSGCRLTSLR	SEKDGAATGV DAICIHRLDP	(SEQ ID NO: 13)
46	ERVLQGLLGP	MFKNTSVGLL	YSGCRLTLLR	PEKNGAATGM DAICSHRLDP	(SEQ ID NO: 14)
33	ERVLQGLLGP	LFKNSSVGPL	YSGCRLISLR	SEKDGAATGV DAICTHHLNP	(SEQ ID NO: 15)
15	ERVLQGLLRP	LFKSTSAGPL	YSGCRLTLLR	PEKHGAATGV DAICTLRLDP	(SEQ ID NO: 16)
35	ERVLQGLLKP	LFKSTSVGPL	YSGCRLTLLR	PEKRGAATGV DTICTHRLDP	(SEQ ID NO: 17)
111	ERVLQGLLTP	LFKNTSVGPL	YSGCRLTLLR	PEKQEAATGV DTICTHRVDP	(SEQ ID NO: 18)
42	ERVLQGLLKP	LFKNTSVGPL	YSGCRLTLLR	PEKHEAATGV DTICTHRLDP	(SEQ ID NO: 19)
116	ERVLQGLLSP	IFKNSSVGPL	YSGCRLTSLR	PEKDGAATGM DAVCLYHPNP	(SEQ ID NO: 20)
23	ERVLQGLLRP	LFKNTSIGPL	YSSCRLTLLR	PEKDKAATRV DAICTHHPDP	(SEQ ID NO: 21)

	51			100
12	KSPRLDREQL	YWELSQLTHN	ITELGPYALD	NDSLFVNGFT HRSSVSTTST
34	KSPGLDRERL	YWKLSQLTHG	ITELGPYTLD	RHSLYVNGFT HQSSMTTTRT
32	KSPGLNREQL	YWELSKLTND	IEELGPYTLD	RNSLYVNGFT HQSSVSTTST
46	KSPGLNREQL	YWELSQLTHG	IKELGPYTLD	RNSLYVNGFT HRSSVAPTST
33	QSPGLDREQL	YWQLSQMTNG	IKELGPYTLD	RNSLYVNGFT HRSSGLTTST
15	TGPGLDRERL	YWELSQLTNS	VTELGPYTLD	RDSLYVNGFT HRSSVPTTSI
35	LNPGLDREQL	YWELSKLTRG	IIELGPYTLD	RDSLYVNGFT HRSSVPTTSI
111	IGPGLDRERL	YWELSQLTNS	ITELGPYTLD	RDSLYVDGFN PWSSVPTTST
42	LNPGLDREQL	YWELSKLTRG	IIELGPYLLD	RGSLYVNGFT HRNFVPITST
116	KRPGLDREQL	YWELSQLTHN	ITELGPYSLD	RDSLYVNGFT HQNSVPTTST
23	QSPGLNREQL	YWELSQLTHG	ITELGPYTLD	RDSLYVDGFT HWSPIPTTST

	101			150
12	PGTPTVYLGA	SKTPASIFGP	S..AASPLLI	PFT~~~~~~~ ~~~~~~~~~~
34	PDTSTMHLAT	SRTPASLSGP	T..TASPLLI	PF~~~~~~~~ ~~~~~~~~~~
32	PGTSTVDLRT	SGTPSSLSSP	TIMAAGPLLI	PF~~~~~~~~ ~~~~~~~~~~
46	PGTSTVDLGT	SGTPSSLPSP	T..TAVPLLI	PF~~~~~~~~ ~~~~~~~~~~
33	PWTSTVDLGT	SGTPSPVPSP	T..TAGPFLI	PF~~~~~~~~ ~~~~~~~~~~
15	PGTSAVHLET	SGTPASLPGH	T..APGPLLI	PF~~~~~~~~ ~~~~~~~~~~
35	PGTSAVHLET	SGTPASLPGH	I..VPGPLLI	PF~~~~~~~~ ~~~~~~~~~~
111	PGTSTVHLAT	SGTPSPLPGH	T..APVPLLI	PFT~~~~~~~ ~~~~~~~~~~
42	PGTSTVHLGT	SETPSSLPRP	I..VPGPLLV	PFT~~~~~~~ ~~~~~~~~~~
116	PGTSTVYWAT	TGTPSSFPGH	T..EPGPLLI	PF~~~~~~~~ ~~~~~~~~~~
23	PGTSIVNLGT	SGIPPSLPET	T..ATGPLLI	PFT~~~~~~~ ~~~~~~~~~~

TABLE 3-continued

Amino Acid Sequence for a 400 bp Repeat in the CA125 Molecule (SEQ ID NO: 11 thru SEQ ID NO: 21)

	151 170
12	~~~~~~~~~~ ~~~~~~~~~~
34	~~~~~~~~~~ ~~~~~~~~~~
32	~~~~~~~~~~ ~~~~~~~~~~
46	~~~~~~~~~~ ~~~~~~~~~~
33	~~~~~~~~~~ ~~~~~~~~~~
15	~~~~~~~~~~ ~~~~~~~~~~
35	~~~~~~~~~~ ~~~~~~~~~~
111	~~~~~~~~~~ ~~~~~~~~~~
42	~~~~~~~~~~ ~~~~~~~~~~
116	~~~~~~~~~~ ~~~~~~~~~~
23	~~~~~~~~~~ ~~~~~~~~~~

TABLE 4

Amino Acid Sequence for a 800 bp Repeat in the CA125 Molecule
(SEQ ID NO: 22 thru SEQ ID NO: 35)

	1				50	
79	ERVLQGLLKP	LFRNSSLEYL	YSGCRLASLR	PEKDSSAMAV	DAICTHRPDP	(SEQ ID NO: 22)
811	ERVLQGLLKP	LFRNSSLEYL	YSGCRLASLR	PEKDSSAMAV	DAICTHRPDP	(SEQ ID NO: 23)
21	ERVLQGLLKP	LFKSTSVGPL	YSGCRLTLLR	PEKRGAATGV	DTICTHRLDP	(SEQ ID NO: 24)
89	ERVLQGLLKP	LFKSTSVGPL	YSGCRLTLLR	PEKRGAATGV	DTICTHRLDP	(SEQ ID NO: 25)
85	ERVLQGLLKP	LFKSTSVGPL	YSGCRLTLLR	PEKRGAATGV	DTICTHRLDP	(SEQ ID NO: 26)
712	ERVLQGLLKP	LFKSTSVGPL	YSGCRLTLLR	PEKRGAATGV	DTICTHRLDP	(SEQ ID NO: 27)
86	ERVLQGLLKP	LFKSTSVGPL	YSGCRLTLLR	PEKHGAATGV	DAICTLRLDP	(SEQ ID NO: 28)
87	ERVLQGLLTP	LFKNTSVGPL	YSGCRLTLLR	PEKQEAATGV	DTICTHRVDP	(SEQ ID NO: 29)
810	ERVLQGLLRP	LFKNTSIGPL	YSSCRLTLLR	PEKDKAATRV	DAICTHHPDP	(SEQ ID NO: 30)
83	ERVLQGLLRP	VFKNTSVGPL	YSGCRLTLLR	PKKDGAATKV	DAICTYRPDP	(SEQ ID NO: 31)
81	ERVLQGLLGP	MFKNTSVGLL	YSGCRLTLLR	PKKDGAATKV	DAICTYRPDP	(SEQ ID NO: 32)
44	ERVLQGLLKP	LFKSTSVGPL	YSGCRLTLLR	PEKDGAATGM	DAVCLYHPNP	(SEQ ID NO: 33)
812	ERVLQGLLSP	ISKNSSVGPL	YSGCRLTSLR	PEKDGAATGM	DAVCLYHPNP	(SEQ ID NO: 34)
76	ERVLQGLLSP	IFKNSSVGSL	YSGCRLTLLR	PEKDGAATRV	DAVCTHRPDP	(SEQ ID NO: 35)

	51				100
79	EDLGLDRERL	YWELSNLTNG	IQELGPYTLD	RNSLYVNGFT	HRSSMPTTST
811	EDLGLDRERL	YWELSNLTNG	IQELGPYTLD	RNSLYVNGFT	HRSSGLTTST
21	LNPGLDREQL	YWELSKLTRG	IIELGPYLLD	RGSLYVNGFT	HRTSVPTTST
89	LNPGLDREQL	YWELSKLTRG	IIELGPYLLD	RGSLYVNGFT	HRNFVPITST
85	LNPGLDREQL	YWELSKLTRG	IIELGPYLLD	RGSLYVNGFS	RQSSMTTTRT
712	LNPGLDREQL	YWELSKLTRG	IIELGPYLLD	RDSLYVNGFT	HRSSVPTTSI
86	TGPGLDRERL	YWELSQLTNS	VTELGPYTLD	RDSLYVNGFT	HRSSVPTTSI
87	IGPGLDRERL	YWELSQLTNS	ITELGPYTLD	RDSLYVNGFN	PWSSVPTTST
810	QSPGLNREQL	YWELSQLTHG	ITELGPYTLD	RDSLYVDGFT	HWSPIPTTST
83	KSPGLDREQL	YWELSQLTHS	ITELGPYTLD	RDSLYVNGFT	QRSSVPTTSI
81	KSPGLDREQL	YWELSQLTHS	ITELGPYTLD	RDSLYVNGFT	QRSSVPTTSI
44	KRPGLDREQL	YCELSQLTHD	ITELGPYSLD	RDSLYVNGFT	HQNSVPTTST
812	KRPGLDREQL	YWELSQLTHN	ITELGPYSLD	RDSLYVNGFT	HQNSVPTTST
76	KSPGLDRERL	YWKLSQLTHG	ITELGPYTLD	RHSLYVNGFT	HQSSMTTTRT

	101				150
79	PGTSTVDVGT	SGTPSSSPSP	TTAGPLLMPF	TLNFTITNLQ	YEEDMRRTGS
811	PWTSTVDLGT	SGTPSPVPSP	TTAGPLLIPF	TLNFTITNLQ	YEENMGHPGS
21	PGTSTVDLGT	SGTPFSLPSP	ATAGPLLVLF	TLNFTITNLK	YEEDMHRPGS
89	PGTSTVHLGT	SETPSSLPRP	IVPGPLLIPF	TINFTITNLR	YEENMHHPGS
85	PDTSTMHLAT	SRTPASLSGP	TTASPLLIPF	TLNFTITNLQ	YEENMGHPGS
712	PGTSAVHLET	FGTPASLHGH	TAPGPVLVPF	TLNFTITNLQ	YEEDMRHPGS
86	PGTSAVHLET	SGTPASLPGH	TAPGPLLVPF	TLNFTITNLQ	YEEDMRHPGS
87	PGTSTVHLAT	SGTPSSLPGH	TAPVPLLIPF	TLNFTITNLH	YEENMQHPGS
810	PGTSIVNLGT	SGIPPSLPET	TATGPLLIPF	TPNFTITNLQ	YEEDMRRTGS
83	PGTPTVDLGT	SGTPVSKPGP	SAASPLLVPF	TLNFTITNLQ	YEEDMHRPGS
81	PGTPTVDLGT	SGTPVSKPGP	SAASPLLIPF	TINFTITNLR	YEENMGHPGS
44	PGTSTVYWAT	TGTPSSFPGH	TEPGPLLIPF	TFNFTITNLH	YEENMQHPGS
812	PGTSTVYWAT	TGTPSSFPGH	TEPGPLLIPF	TVNFTITNLR	YEENMHHPGS
76	PDTSTMHLAT	SRTPASLSGP	TTASPLLVLF	TINFTITNQR	YEENMHHPGS

TABLE 4-continued

Amino Acid Sequence for a 800 bp Repeat in the CA125 Molecule
(SEQ ID NO: 22 thru SEQ ID NO: 35)

```
      151                                                        200
 79   RKFNTMERVL QGLLSPIFKN SSVGPLYSGC RLTSLRPEKD GAATGMDAVC
811   RKFNIMERVL QGLLMPLFKN TSVSSLYSGC RLTLLRPEKD GAATRVDAVC
 21   RKFNTTERVL QTLLGPMFKN TSVGLLYSGC RLTLLRSEKD GAATGVDAIC

 89   RKFNIMERVL QGLLGPLFKN SSVGPLYSGC RLISLRSEKD GAATGVDAIC
 85   RKFNIMERVL QGLLNPIFKN SSVGPLYSGC RLTSLKPEKD GAATGMDAVC
712   RKFNTTERVL QGLLKPLFKS TSVGPLYSGC RLTLLRPEKR GAATGVDTIC
 86   RKFNTTERVL QGLLKPLFKS TSVGPLYSGC RLTLLRPEKR GAATGVDTIC
 87   RKFNTTERVL QGLLKPLFKS TSVGPLYSGC RLTLLRPEKH GAATGVDAIC
810   RKFNTMERVL QGLLSPIFKN SSVGPLYSGC RLTSLRPEKD GAATGMDAVC
 83   RKFNATERVL QGLLSPIFKN SSVGPLYSGC RLTSLRPEKD GAATGMDAVC
 81   RKFNIMERVL QGLLKPLFKN TSVGPLYSGC RLTLLRPKKD GAATGVDAIC
 44   RKFNTTERVL QGLLKPLFKN TSVGPLYSGC RLTLLRPEKH EAATGVDTIC
812   RKFNTTERVL QGLLRPVFKN TSVGPLYSGC RLTLLRPKKD GAATKVDAIC
 76   RKFNTTERVL QGLLRPVFKN TSVGPLYSGC RLTLLRPKKD GAATKVDAIC

      201                                                        250
 79   LYHPNPKRPG LDREQLYWEL SQLTHNITEL GPYSLDRDSL YVNGFTHQNS
811   TQRPDPKSPG LDRERLYWKL SQLTHGITEL GPYTLDRHSL YVNGLTHQSS
 21   THRLDPKSPG VDREQLYWEL SQLTNGIKEL GPYTLDRNSL YVNGFTHWIP
 89   THHLNPQSPG LDREQLYWQL SQMTNGIKEL GPYTLDRNSL YVNGFTHRSS
 85   LYHPNPKRPG LDREQLYWEL SQLTHGIKEL GPYTLDRNSL YVNGFTHRSS
712   THRLDPLNPG LDREQLYWEL SKLTRGIIEL GPYLLDRGSL YVNGFTHRNF
 86   THRLDPLNPG LDREQLYWEL SKLTRGIIEL GPYLLDRGSL YVNGFTHRNF
 87   THRLDPKSPG VDREQLYWEL SQLTNGIKEL GPYTLDRNSL YVNGFTHWIP
810   LYHPNPKRPG LDREQLY~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 83   LYHPNPKRPG LDREQLYWEL SQLTHNITEL GPYSLDRDSL YVNGFTHQSS
 81   THRLDPKSPG LNREQLYWEL SKLTNDIEEL GPYTLDRNSL YVNGFTHQSS
 44   THRVDPIGPG LDRERLYWEL SQLTNSIHEL GPYTLDRDSL YVNGFNPRSS
812   TYRPDPKSPG LDREQLYWEL SKLTNDIEEL GPYTLDRNSL YVNGFTHQSS
 76   TYRPDPKSPG LDREQLYWEL SQLTHSITEL GPYTQDRDSL YVNGFTHRSS

      251                                          288
 79   VPTTSTPGTS TVYWATTGTP SSFPGHT..E PGPL~~~~
811   MTTTRTPDTS TMHLATSRTP ASLSGPT..T ASPLLIPF
 21   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~
 89   GLTTSTPWTS TVDLGTSGTP SPVPSPT..T AGPLLIPF
 85   VAPTSTPGTS TVDLGTSGTP SSLPSPT..T AVPLLIPF
712   VPITSTPGTS TVHLGTSETP SSLPRPI..V PGPLLIPF
 86   VPITSTPGTS TVHLGTSETP SSLPRPI..V PGPLLIPF
 87   VPTSSTPGTS TVDLG.SGTP SSLPSPT..T AGPL~~~~
810   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~
 83   MTTTRTPDTS TMHLATSRTP ASLSGPT..T ASPLLIPF
 81   VSTTSTPGTS TVDLRTSGTP SSLSSPTIMA AGPLLIPF
 44   VPTTSTPGTS TVHLATSGTP SSLPGHT..A PVPLLI~~
812   VSTTSTPGTS TVDLRTSGTP SSLSSPTIMA AGPLLIPF
 76   VPTTSIPGTS AVHLETSGTP ASLP~~~~~~ ~~~~~~~~
```

TABLE 5

Amino Acid Sequence for a 1200 bp Repeat in the CA125 Molecule (SEQ ID NO: 36 thru SEQ ID NO: 46)

	1				50	
910	ERVLQGLLGP	MFKNTSVGLL	YSGCRLTLLR	PEKRGAATGV	DTICTHRLDP	(SEQ ID NO: 36)
99	ERVLHGLLTP	LFKNTRVGPL	YSGCRLTLLR	PEKQEAATGV	DTICTHRVDP	(SEQ ID NO: 37)
112	~~~~~~~~~~	~~~~~~~GPL	YSGCRLTSLR	PEKDGAATGM	DAVCLYHPNP	(SEQ ID NO: 38)
95	ERVLQGPLSP	IFKNSSVGPL	YSGCRLTSLR	PEKDGAATGM	DAVCLYHPNP	(SEQ ID NO: 39)
71	~~~~~~~~~~	~~~~TSVGPL	YSGCRLTLLR	SEKDGAATGV	DAIYTHRLDP	(SEQ ID NO: 40)
78	~~~~~~~~~~	~~~~~~~~~~	~~~~~~TLLR	PKKDGVATGV	DAICTHRLDP	(SEQ ID NO: 41)
115	ERVLQGLLKP	LFKSTSVGPL	YSGCRLTLLR	PEKDGVATRV	DAICTHRPDP	(SEQ ID NO: 42)
91	ERVLQGLLKP	LFRNSSLEYL	YSGCRLASLR	PEKDSSAMAV	DAICTHRPDP	(SEQ ID NO: 43)
92	ERVLQGLLKP	LFKSTSVGPL	YSGCRLTLLR	PEKRGAATGV	DTICTHRLDP	(SEQ ID NO: 44)
113	ERVLQGLLGP	MFKNTSVGLL	YSGCRLTLLR	PEKNGAATGM	DAICSHRLDP	(SEQ ID NO: 45)
711	ERVLQGLLKP	LFKSTSVGPL	YSGCRLTLLR	PEKHGAATGV	DAICTLRLDP	(SEQ ID NO: 46)

	51				100
910	LNPGLDREQL	YWELSKLTRG	IIELGPYLLD	RGSLYVNGFT	HRNFVPITST
99	IGPGLDRERL	YWELSQLTNS	ITELGPYTLD	RDSLYVNGFN	PWSSVPTTST
112	KRPGLDREQL	YWELSQLTHN	ITELGPYSLD	RDSLYVNGFT	HQNSVPTTST
95	KRPGLDREQL	YWELSQLTHN	ITELGPYSLD	RDSLYVNGFT	HQNSVPTTST
71	KSPGVDREQL	YWELSQLTNG	IKELGPYTLD	RNSLYVNGFT	HQTSAPNTST
78	KSPGLNREQL	YWELSKLTND	IEELGPYTLD	RNSLYVNGFT	HQSSVSTTST
115	KIPGLDRQQL	YWELSQLTHS	ITELGPYTLD	RDSLYVNGFT	QRSSVPTTST
91	EDLGLDRERL	YWELSNLTNG	IQELGPYTLD	RNSLYVNGFT	HRSSMPTTST
92	LNPGLDREQL	YWELSKLTRG	IIELGPYLLD	RGSLYVNGFT	HRNFVPITST
113	KSPGLNREQL	YWELSQLTHG	IKELGPYTLD	RNSLYVNGFT	HRSSVAPTST
711	TGPGLDRERL	YWELSQLTNS	VTELGPYTLD	RDSLYVNGFT	HRSSVPTTSI

	101				150
910	PGTSTVHLGT	SETPSSLPRP	IV..PGPLLV	PFTLNFTITN	LQYEEAMRHP
99	PGTSTVHLAT	SGTPSSLPGH	TA..PVPLLI	PFTLNFTITN	LHYEENMQHP
112	PGTSTVYWAT	TGTPSSFPGH	T..EPGPLLI	PFTLNFTITN	LQYEENMGHP
95	PGTSTVYWAT	TGTPSSFPGH	T..EPGPLLI	PFTLNFTITN	LQYEENMGHP
71	PGTSTVDLGT	SGTPSSLPSP	T..SAGPLLI	PFTINFTITN	LRYEENMHHP
78	PGTSTVDLRT	SGTPSSLSSP	TIMAAGPLLI	PFTINFTITN	LRYEENMHHP
115	PGTFTVQPET	SETPSSLPGP	T..ATGPVLL	PFTLNFTIIN	LQYEEDMHRP
91	PGTSTVDVGT	SGTPSSSPSP	T..TAGPLLM	PFTLNFTITN	LQYEEDMRRT
92	PGTSTVHLGT	SETPSSLPRP	IV..PGPLLI	PFTLNFTITN	LQYEENMGHP
113	PGTSTVDLGT	SGTPSSLPSP	T..TAVPLLI	PFTLNFTITN	LKYEEDMHCP
711	PGTSAVHLET	SGTPASLPGH	T..APGPLLI	PFTLNFTITN	LHYEENMQHP

	151				200
910	GSRKFNTTER	VLQGLLRPLF	KNTSVSSLYS	GCRLTLLRPE	KDGAATRVDA
99	GSRKFNTTER	VLQGLLKPLF	KNTSVGPLYS	GCRLTLFKPE	KHEAATGVDA
112	GSRKFNITES	VLQGLLTPLF	KNSSVGPLYS	GCRLISLRSE	KDGAATGVDA
95	GSRKFNITER	VLQGLLNPIF	KNSSVGPLYS	GCRLTSLRPE	KDGAATGMDA
71	GSRKFNTMER	VLQGLLKPLF	KSTSVGPLYS	GCRLTLLRPE	KDGVATRVDA
78	GSRKFNTMER	VLQGLLMPLF	KNTSVSSLYS	GCRLTLLRPE	KDGAATRVDA
115	GSRKFNTTER	VLQGLLMPLF	KNTSVGPLYS	GCRLTLLRPE	KQEAATGVDT
91	GSRKFNTMES	VLQGLLKPLF	KNTSVGPLYS	GCRLTLLRPK	KDGAATGVDA
92	GSRKFNITER	VLQGLLKPLF	RNSSLEYLYS	GCRLTSLRPE	KDSSTMAVDA

TABLE 5-continued

Amino Acid Sequence for a 1200 bp Repeat in the CA125 Molecule
(SEQ ID NO: 36 thru SEQ ID NO: 46)

```
113  GSRKFNTTER VLQSLFGPMF KNTSVGPLYS GCRLTLFRSE KDGAATGVDA
711  GSRKFNTMER VLQGCLVPCS RNTNVGLLYS GCRLTLLXXX XXXXXXXXXX

                                                        250
     201
910  ACTYRPDPKS PGLDREQLYW ELSQLTHSIT ELGPYTLDRV SLYVNGFNPR
 99  ICTLRLDPTG PGLDRERLYW ELSQLTNSVT ELGPYTLDRD SLYVNGFTHR
112  ICTHHLNPQS PGLDREQLYW QLSQMTNGIK ELGPYTLDRD SLYVNGFTHR
 95  VCLYHPNPKR PGLDREQLYC ELSQLTHNIT ELGPYSLDRD SLYVNGFTHQ
 71  ICTHRPDPKI PGLDRQQLYW ELSQLTHSIT ELGPYTLDRD SLYVNGFTQR
 78  VCTHRPDPKS PGLDRERLYW KLSQLTHGIT ELGPYTLDRN SLYVNGFTHR
115  ICTHRLDPSE PGLDREQLYW ELSQLTNSIT ELGPYTLDRD SLYVNGFTHS
 91  ICTHRLDPKS PGLNREQLYW ELSKLTNDIE EVGPYTLDRN SLYVNGFTHR
 92  ICTHRPDPED LGLDRERLYW ELSNLTNGIQ ELGPYTLDRN SLYVNGFTHR
113  ICTHRLDPKS PGVDREQLYW ELSQLTNGIK ELGPYTLDRN SLYVNGFTHQ
711  XXXXXXXXXX XXXXXXXXXX XXXXXXXXXX XXGPYTLDRN SLYVNGFTHR

                                                        300
     251
910  SSV.PTTSTP GTSTVHLATS GTPSSLPGHT APVPLLIPFT LNFTITNLQY
 99  SSV.PTTSIP GTSAVHLETS GTPASLPGHT APGPLLIPFT LNFTITNLQY
112  SL.GLTTSTP WTSTVDLGTS GTPSPVPSPT TAGPLLIPFT LNFTITNLQY
 95  NS.VPTTSTP GTSTVYWATT GTPSSFPGHT EPGPLLIPFT LNFTITNLQY
 71  SSV.PTTSTP GTFTVQPETS ETPSSLPGPT ATGPVLLPFT LNFTIINLQY
 78  SSM.PTTSTP GTSTVDVGTS GTPSSSPSPT TAGPLLMPFT LNFTITNLQY
115  GVLCPPPSIL GIFTVQPETF ETPSSLPGPT ATGPVLLPFT LNFTIINLQY
 91  SFVAP.TSTL GTSTVDLGTS GTPSSLPSPT TGVPLLIPFT LNFTITNLQY
 92  SFM.PTTSTL GTSTVDVGTS GTPSSSPSPT TAGPLLMPFT LNFTITNLQY
113  TS.APNTSTP GTSTVDLGTS GTPSSLPSPT SAGPLLVPFT LNFTITNLQY
711  SSVAP.TSTP GTSTVDLGTS GTPSSLPSPT TV.PLLVPFT LNFTITNLQY

                                                        350
     301
910  EEDMRHPGSR KFNTMERVLQ GLLRPLFKNT SIGPLYSSCR LTLLRPEKDK
 99  EEDMRRTGSR KFNTMERVLQ GLLKPLFKST SVGPLYSGCR LTLLRPEKRG
112  EENMGHPGSR KFNIMERVLQ GLLRPVFKNT SVGPLYSGCR LTLLRPKKDG
 95  EEDMRRTGSR KFNTMERVLQ GLLKPLFKST SVGPLYSGCR LTLLRPEKHG
 71  EEDMHRPGSR KFNTTERVLQ GLLKPLFKST SVGPLYSGCR LTLLRPEKHG
 78  EEDMRRTGSR KFNTMERVLQ GLLKPLFKST SVGPLYSGCR LTLLRPEKHG
115  EEDMHRPGSR KFNTTERVLQ GLLMPLFKNT SVGPLYSGCR LTLLRPEKQE
 91  EENMGHPGSR KFNIMERVLQ GLLMPLFKNT SVSSLYSGCR LTLLRPEKDG
 92  EEDMRRTGSR KFNTMESVLQ GLLKPLFKNT SVGPLYSGCR LTLLRPKKDG
113  EEDMRRTGSR KFNTMESVLQ GLLKPLFKNT SVGPLYSGCR LTLLRPEKDG
711  GEDMRHPGSR KFNTTERVLQ GLLGPLFKNS SVGPLYSGCR LISLRSEKDG

                                                        400
     351
910  AATRVDAICT HHPDPQSPGL NREQLYWELS QLTHGITEL- ----------
 99  AATGVDTICT HRLDPLNPGL DREQLYWELS KLTRGIIELG PYLLDRGSLY
112  AATKVDAICT YRPDPKSPGL DREQLYWELS QLTHSITELG PYTLDRDSLY
 95  AATGVDAICT LRLDPTGPGL DRERLYWELS QLTNSVTELG PYTLDRDSLY
 71  AATGVDAICT LRLDPTGPGL DRERLYWELS QLTNSITELG PYTLDRDSLY
 78  AATGVDAICT LRLDPTGPGL DRERLYWELS QLTNSVTELG PYTLDRDSLY
```

TABLE 5-continued

Amino Acid Sequence for a 1200 bp Repeat in the CA125 Molecule
(SEQ ID NO: 36 thru SEQ ID NO: 46)

```
115  AATGVDTICT HRVDPIGPGL DRERLYWELS QLTNSITELG PYTLDRDSLY
 91  AATRVVAVCT HRPDPKSPGL DRERLYWKLS QLTHGITELG PYTLDRHSLY
 92  AATGVDAICT HRLDPKSPGL NREQLYWELS KLTNDIEELG PYTLDRNSLY
113  AATGVDAICT HRLDPKSPGL NREQLYWELS KL~~~~~~~~ ~~~~~~~~~~
711  AATGVDAICT HHLNPQSPGL DREQLYWQLS QVTNGIKELG PYTLDRNSLY

     401                                                 447

910  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~
 99  VNGFTHRNFV PITSTPGTST VHLGTSEIHP SLPRPI..VP GPL~~~~
112  VNGFTQRSSV PTTSIPGTPT VDLGTSGTPV SKPGPS..AA SP~~~~~
 95  VNGFTHRSSV PTTSIPGTSA VHLETSGTPA SLPGHT..AP GPLL~~~
 71  VNGFNPWSSV PTTSTPGTST VHLATSGTPS SLPGHT..AP VPL~~~~
 78  VNGFTHRSSV PTTSIPGTSA VHLETSGTPA SLPGHT..AP GPLLIPF
115  VNGFNPWSSV PTTSTPGTST VHLATSGTPS SLPGHT..AP VPLLIPF
 91  VNGFTHQSSM TTTRTPDTST MHLATSRTPA SLSGPT..TA SPLLIPF
 92  VNGFTHQSSV STTSTPGTST VDPRTSGTPS SLSSPTIMAA GPLLI~~
113  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~
711  VNGFTHRSSG LTTSTPWTST VDLGTSGTPS PVPSPT..TA GPLLI~~
```

TABLE 6

Amino Acid Sequence for a 9 Repeat Structure in the CA125 Molecule (SEQ ID NO: 47)

ERVLQGLLKP LFRNSSLEYL YSGCRLASLR PEKDSSAMAV DAICTHRPDP
EDLGLDRERL YWELSNLTNG IQELGPYTLD RNSLYVNGFT HRSSMPTTST
PGTSTVDVGT SGTPSSSPSP TTAGPLLMPF TLNFTITNLQ YEEDMRRTGS
RKFNTMERVL QGPLSPIFKN SSVGPLYSGC RLTSLRPEKD GAATGM DAV
CLYHPNPKRP GLDREQLYWE LSQLTHNITE LGPYSLDRDS LYVNGFTHQN
SVPTTSTPGT STVYWATTGT PSSFPGHTEP GPLLIPFTLN FTITNLQYEE
NMGHPGSRKF NITERVLQGL LNPIFKNSSV GPLYSGCRLT SLRPEKDGAA
TGMDAVCLYH PNPKRPGLDR EQLYCELSQL THNITELGPY SLDRDSLYVN
GFTHQNSVPT TSTPGTSTVY WATTGTPSSF PGHTEPGPLL IPFTLNFTIT
NLQYEEDMRR TGSRKFNTME RVLQGLLKPL FKSTSVGPLY SGCRLTLLRP
EKHGAATGVD AICTLRLDPT GPGLDRERLY WELSQLTNSV TELGPYTLDR
DSLYVNGFTH RSSVPTTSIP GTSAVHLETS GTPASLPGHT APGPLLVPFT
LNFTITNLQY EEDMRHPGSR KFNTTERVLQ GLLKPLFKST SVGPLYSGCR
LTLLRPEKRG AATGVDTICT HRLDPLNPGL DREQLYWELS KLTRGIIELG
PYLLDRGSLY VNGFTHRNFV PITSTPGTST VHLGTSETPS SLPRPIVPGP
LLIPFTLNFT ITNLQYEENM GHPGSRKFNI TERVLQGLLK PLFRNSSLEY
LYSGCRLASL RPEKDSSAMA VDAICTHRPD PEDLGLDRER LYWELSNLTN
GIQELGPYTL DRNSLYVNGF THRSSMPTTS TPGTSTVDVG TSGTPSSSPS
PTTAGPLLMP FTLNFTITNL QYEEDMRRTG SRKFNTMESV LQGLLKPLFK
NTSVGPLYSG CRLTLLRPKK DGAATGVDAI CTHRLDPKSP GLNREQLYWE
LSKLTNDIEE VGPYTLDRNS LYVNGFTHRS FVAPTSTLGT STVDLGTSGT
PSSLPSPTTG VPLLIPFTLN FTITNLQYEE NMGHPGSRKF NIMERVLQGL
LSPIFKNSSV GSLYSGCRLT LLRPEKDGAA TRVDAVCTHR PDPKSPGLDR
ERLYWKLSQL THGIIELGPY TLDRHSFYVN GFTHQSSMTT TRTPDTSTMH
LATSRTPASL SGPTTASPLL VLFTINFTIT NQRYEENMHH PGSRKFNTTE
RVLQGLLRPV FKNTSVGPLY SGCRLTLLRP KKDGAATKVD AICTYRPDPK
SPGLDREQLY WELSQLTHSI TELGPYTQDR DSLYVNGFTH RSSVPTTSIP
GTSAVHLETS GTPASLP

TABLE 7

cDNA Genbank Accession # AK024365 Encompasses Repeat Sequences (Repeats 1 & 2) Homologous to Two Repeats Shown in Table 6 (SEQ ID NO: 48)

MPLFKNTSVS SLYSGCRLTL LRPEKDGAAT RVDAVCTHRP DPKSPGLDRE
RLYWKLSQLT HGIIELGPYT LDRHSFYVNG FTHQSSMTTT RTPDTSTMHL
ATSRTPASLS GPTTASPLLV LFTINFTITN QRYEENMHHP GSRKFNTTER
VLQGLLRPVF KNTSVGPLYS GCRLTLLRPK KDGAATKVDA ICTYRPDPKS
PGLDREQLYW ELSQLTHSIT ELGPYTQDRD SLYVNGFTHR SSVPTTSIPG
TSAVHLETSG TPASLPGPSA ASPLLVLFTL NFTITNLRYE ENMQHPGSRK
FNTTERVLQG LLRSLFKSTS VGPLYSGCRL TLLRPEKDGT ATGVDAICTH
HPDPKSPRLD REQLYWELSQ LTHNITELGH YALDNDSLFV NGFTHRSSVS
TTSTPGTPTV YLGASKTPAS IFGPSAASHL LILFTLNFTI TNLRYEENMW
PGSRKFNTTE RVLQGLLRPL FKNTSVGPLY SGSRLTLLRP EKDGEATGVD
AICTHRPDPT GPGLDREQLY LELSQLTHSI TELGPYTLDR DSLYVNGFTH
RSSVPTTSTG VVSEEPFTLN FTINNLRYMA DMGQPGSLKF NITDNVMKHL
LSPLFQRSSL GARYTGCRVI ALRSVKNGAE TRVDLLCTYL QPLSGPGLPI
KQVFHELSQQ THGITRLGPY SLDKDSLYLN GYNEPGLDEP PTTPKPATTF
LPPLSEATTA MGYHLKTLTL NFTISNLQYS PDMGKGSATF NSTEGVLQHL
LRPLFQKSSM GPFYLGCQLI SLRPEKDGAA TGVDTTCTYH PDPVGPGLDI
QQLYWELSQL THGVTQLGFY VLDRDSLFIN GYAPQNLSIR GEYQINFHIV
NWNLSNPDPT SSEYITLLRD IQDKVTTLYK GSQLHDTFRF CLVTNLTMDS
VLVTVKALFS SNLDPSLVEQ VFLDKTLNAS FHWLGSTYQL VDIHVTEMES
SVYQPTSSSS TQHFYLNFTI TNLPYSQDKA QPGTTNYQRN KRNIEDALNQ
LFRNSSIKSY FSDCQVSTFR SVPNRHHTGV DSLCNFSPLA RRVDRVAIYE
EFLRMTRNGT QLQNFTLDRS SVLVDGYSPN RNEPLTGNSD LPFWAVILIG
LAGLLGLITC LICGVLVTTR RRKKEGEYNV QQQCPGYYQS HLDLEDLQ

TABLE 8

Complete DNA Sequence for 13 Repeats including the Carboxy Terminus of CA125 (SEQ ID NO: 49)

```
   1  GAGAGGGTTC TGCAGGGTCT GCTCAAACCC TTGTTCAGGA ATAGCAGTCT
  51  GGAATACCTC TATTCAGGCT GCAGACTAGC CTCACTCAGG CCAGAGAAGG
 101  ATAGCTCAGC CATGGCAGTG GATGCCATCT GCACACATCG CCCTGACCCT
 151  GAAGACCTCG GACTGGACAG AGAGCGACTG TACTGGGAGC TGAGCAATCT
 201  GACAAATGGC ATCCAGGAGC TGGGCCCCTA CACCCTGGAC CGGAACAGTC
 251  TCTATGTCAA TGGTTTCACC CATCGAAGCT CTATGCCCAC CACCAGCACT
 301  CCTGGGACCT CCACAGTGGA TGTGGGAACC TCAGGGACTC CATCCTCCAG
 351  CCCCAGCCCC ACGACTGCTG GCCCTCTCCT GATGCCGTTC ACCCTCAACT
 401  TCACCATCAC CAACCTGCAG TACGAGGAGG ACATGCGTCG CACTGGCTCC
 451  AGGAAGTTCA ACACCATGGA GAGGGTTCTG CAGGGTCCGC TTAGTCCCAT
 501  ATTCAAGAAC TCCAGTGTTG GCCCTCTGTA CTCTGGCTGC AGACTGACCT
 551  CTCTCAGGCC CGAGAAGGAT GGGGCAGCAA CTGGAATGGA TGCTGTCTGC
 601  CTCTACCACC CTAATCCCAA AAGACCTGGG CTGGACAGAG AGCAGCTGTA
 651  CTGGGAGCTA AGCCAGCTGA CCCACAACAT CACTGAGCTG GGCCCCTACA
 701  GCCTGGACAG GGACAGTCTC TATGTCAATG GTTTCACCCA TCAGAACTCT
 751  GTGCCCACCA CCAGTACTCC TGGGACCTCC ACAGTGTACT GGGCAACCAC
 801  TGGGACTCCA TCCTCCTTCC CCGGCCACAC AGAGCCTGGC CCTCTCCTGA
 851  TACCATTCAC GCTCAACTTC ACCATCACTA ACCTACAGTA TGAGGAGAAC
 901  ATGGGTCACC CTGGCTCCAG GAAGTTCAAC ATCACGGAGA GGGTTCTGCA
 951  GGGTCTGCTT AATCCCATTT TCAAGAACTC CAGTGTTGGC CCTCTGTACT
1001  CTGGCTGCAG ACTGACCTCT CTCAGGCCCG AGAAGGATGG GGCAGCAACT
1051  GGAATGGATG CTGTCTGCCT CTACCACCCT AATCCCAAAA GACCTGGGCT
1101  GGACAGAGAG CAGCTGTACT GCGAGCTAAG CCAGCTGACC CACAACATCA
1151  CTGAGCTGGG CCCCTACAGC TTGGACAGGG ACAGTCTTTA TGTCAATGGT
```

TABLE 8-continued

Complete DNA Sequence for 13 Repeats including the Carboxy Terminus of CA125 (SEQ ID NO: 49)	
1201	TTCACCCATC AGAACTCTGT GCCCACCACC AGTACTCCTG GGACCTCCAC
1251	AGTGTACTGG GCAACCACTG GGACTCCATC CTCCTTCCCC GGCCACACAG
1301	AGCCTGGCCC TCTCCTGATA CCATTCACCC TCAACTTCAC CATCACCAAC
1351	CTGCAGTACG AGGAGGACAT GCGTCGCACT GGCTCCAGGA AGTTCAACAC
1401	CATGGAGAGG GTTCTGCAGG GTCTGCTCAA GCCCTTGTTC AAGAGCACCA
1451	GCGTTGGCCC TCTGTACTCT GGCTGCAGAC TGACCTTGCT CAGACCTGAG
1501	AAACATGGGG CAGCCACTGG AGTGGACGCC ATCTGCACCC TCCGCCTTGA
1551	TCCCACTGGT CCTGGACTGG ACAGAGAGCG GCTATACTGG GAGCTGAGCC
1601	AGCTGACCAA CAGCGTTACA GAGCTGGGCC CCTACACCCT GGACAGGGAC
1651	AGTCTCTATG TCAATGGCTT CACCCATCGG AGCTCTGTGC CAACCACCAG
1701	TATTCCTGGG ACCTCTGCAG TGCACCTGGA AACCTCTGGG ACTCCAGCCT
1751	CCCTCCCTGG CCACACAGCC CCTGGCCCTC TCCTGGTGCC ATTCACCCTC
1801	AACTTCACTA TCACCAACCT GCAGTATGAG GAGGACATGC GTCACCCTGG
1851	TTCCAGGAAG TTCAACACCA CGGAGAGAGT CCTGCAGGGT CTGCTCAAGC
1901	CCTTGTTCAA GAGCACCAGT GTTGGCCCTC TGTACTCTGG CTGCAGACTG
1951	ACCTTGCTCA GGCCTGAAAA ACGTGGGGCA GCCACCGGCG TGGACACCAT
2001	CTGCACTCAC CGCCTTGACC CTCTAAACCC TGGACTGGAC AGAGAGCAGC
2051	TATACTGGGA GCTGAGCAAA CTGACCCGTG GCATCATCGA GCTGGGCCCC
2101	TACCTCCTGG ACAGAGGCAG TCTCTATGTC AATGGTTTCA CCCATCGGAA
2151	CTTTGTGCCC ATCACCAGCA CTCCTGGGAC CTCCACAGTA CACCTAGGAA
2201	CCTCTGAAAC TCCATCCTCC CTACCTAGAC CCATAGTGCC TGGCCCTCTC
2251	CTGATACCAT TCACACTCAA CTTCACCATC ACTAACCTAC AGTATGAGGA
2301	GAACATGGGT CACCCTGGCT CCAGGAAGTT CAACATCACG GAGAGGGTTC
2351	TGCAGGGTCT GCTCAAACCC TTGTTCAGGA ATAGCAGTCT GGAATACCTC

TABLE 8-continued

Complete DNA Sequence for 13 Repeats including the Carboxy Terminus of CA125 (SEQ ID NO: 49)	
2401	TATTCAGGCT GCAGACTAAC CTCACTCAGG CCAGAGAAGG ATAGCTCAAC
2451	CATGGCAGTG GATGCCATCT GCACACATCG CCCTGACCCT GAAGACCTCG
2501	GACTGGACAG AGAGCGACTG TACTGGGAGC TGAGCAATCT GACAAATGGC
2551	ATCCAGGAGC TGGGCCCCTA CACCCTGGAC CGGAACAGTC TCTATGTCAA
2601	TGGTTTCACC CATCGAAGCT CTATGCCCAC CACCAGCACT CCTGGGACCT
2651	CCACAGTGGA TGTGGGAACC TCAGGGACTC CATCCTCCAG CCCCAGCCCC
2701	ACGACTGCTG GCCCTCTCCT GATGCCGTTC ACCCTCAACT TCACCATCAC
2751	CAACCTGCAG TACGAGGAGG ACATGCGTCG CACTGGCTCC AGGAAGTTCA
2801	ACACCATGGA GAGTGTCCTG CAGGGTCTGC TCAAGCCCTT GTTCAAGAAC
2851	ACCAGTGTTG GCCCTCTGTA CTCTGGCTGC AGATTGACCT TGCTCAGGCC
2901	CAAGAAAGAT GGGGCAGCCA CTGGAGTGGA TGCCATCTGC ACCCACCGCC
2951	TTGACCCCAA AAGCCCTGGA CTCAACAGGG AGCAGCTGTA CTGGGAGTTA
3001	AGCAAACTGA CCAATGACAT TGAAGAGGTG GGCCCCTACA CCTTGGACAG
3051	GAACAGTCTC TATGTCAATG GTTTCACCCA TCGGAGCTTT GTGGCCCCCA
3101	CCAGCACTCT TGGGACCTCC ACAGTGGACC TTGGGACCTC AGGGACTCCA
3151	TCCTCCCTCC CCAGCCCCAC AACAGGTGTT CCTCTCCTGA TACCATTCAC
3201	ACTCAACTTC ACCATCACTA ACCTACAGTA TGAGGAGAAC ATGGGTCACC
3251	CTGGCTCCAG GAAGTTCAAC ATCATGGAGA GGGTTCTGCA GGGTCTGCTT
3301	ATGCCCTTGT TCAAGAACAC CAGTGTCAGC TCTCTGTACT CTGGTTGCAG
3351	ACTGACCTTG CTCAGGCCTG AGAAGGATGG GGCAGCCACC AGAGTGGTTG
3401	CTGTCTGCAC CCATCGTCCT GACCCCAAAA GCCCTGGACT GGACAGAGAG
3451	CGGCTGTACT GGAAGCTGAG CCAGCTGACC CACGGCATCA CTGAGCTGGG
3501	CCCCTACACC CTGGACAGGC ACAGTCTCTA TGTCAATGGT TTCACCCATC
3551	AGAGCTCTAT GACGACCACC AGAACTCCTG ATACCTCCAC AATGCACCTG

TABLE 8-continued

Complete DNA Sequence for 13 Repeats including the Carboxy Terminus of CA125 (SEQ ID NO: 49)	
3601	GCAACCTCGA GAACTCCAGC CTCCCTGTCT GGACCTACGA CCGCCAGCCC
3651	TCTCCTGATA CCATTCACAA TTAACTTCAC CATCACTAAC CTGCGGTATG
3701	AGGAGAACAT GCATCACCCT GGCTCTAGAA AGTTTAACAC CACGGAGAGA
3751	GTCCTTCAGG GTCTGCTCAG GCCTGTGTTC AAGAACACCA GTGTTGGCCC
3801	TCTGTACTCT GGCTGCAGAC TGACCTTGCT CAGGCCCAAG AAGGATGGGG
3851	CAGCCACCAA AGTGGATGCC ATCTGCACCT ACCGCCCTGA TCCCAAAAGC
3901	CCTGGACTGG ACAGAGAGCA GCTATACTGG GAGCTGAGCC AGCTAACCCA
3951	CAGCATCACT GAGCTGGGCC CCTACACCCT GGACAGGGAC AGTCTCTATG
4001	TCAATGGTTT CACACAGCGG AGCTCTGTGC CCACCACTAG CATTCCTGGG
4051	ACCCCCACAG TGGACCTGGG AACATCTGGG ACTCCAGTTT CTAAACCTGG
4101	TCCCTCGGCT GCCAGCCCTC TCCTGGTGCT ATTCACTCTC AACTTCACCA
4151	TCACCAACCT GCGGTATGAG GAGAACATGC AGCACCCTGG CTCCAGGAAG
4201	TTCAACACCA CGGAGAGGGT CCTTCAGGGC CTGCTCAGGT CCCTGTTCAA
4251	GAGCACCAGT GTTGGCCCTC TGTACTCTGG CTGCAGACTG ACTTTGCTCA
4301	GGCCTGAAAA GGATGGGACA GCCACTGGAG TGGATGCCAT CTGCACCCAC
4351	CACCCTGACC CCAAAAGCCC TAGGCTGGAC AGAGAGCAGC TGTATTGGGA
4401	GCTGAGCCAG CTGACCCACA ATATCACTGA GCTGGGCCAC TATGCCCTGG
4451	ACAACGACAG CCTCTTTGTC AATGGTTTCA CTCATCGGAG CTCTGTGTCC
4501	ACCACCAGCA CTCCTGGGAC CCCCACAGTG TATCTGGGAG CATCTAAGAC
4551	TCCAGCCTCG ATATTTGGCC CTTCAGCTGC CAGCCATCTC CTGATACTAT
4601	TCACCCTCAA CTTCACCATC ACTAACCTGC GGTATGAGGA GAACATGTGG
4651	CCTGGCTCCA GGAAGTTCAA CACTACAGAG AGGGTCCTTC AGGGCCTGCT
4701	AAGGCCCTTG TTCAAGAACA CCAGTGTTGG CCCTCTGTAC TCTGGCTCCA
4751	GGCTGACCTT GCTCAGGCCA GAGAAAGATG GGGAAGCCAC CGGAGTGGAT

TABLE 8-continued

Complete DNA Sequence for 13 Repeats including the Carboxy Terminus of CA125 (SEQ ID NO: 49)

```
4801  GCCATCTGCA CCCACCGCCC TGACCCCACA GGCCCTGGGC TGGACAGAGA
4851  GCAGCTGTAT TTGGAGCTGA GCCAGCTGAC CCACAGCATC ACTGAGCTGG
4901  GCCCCTACAC ACTGGACAGG GACAGTCTCT ATGTCAATGG TTTCACCCAT
4951  CGGAGCTCTG TACCCACCAC CAGCACCGGG GTGGTCAGCG AGGAGCCATT
5001  CACACTGAAC TTCACCATCA ACAACCTGCG CTACATGGCG GACATGGGCC
5051  AACCCGGCTC CCTCAAGTTC AACATCACAG ACAACGTCAT GAAGCACCTG
5101  CTCAGTCCTT TGTTCCAGAG GAGCAGCCTG GGTGCACGGT ACACAGGCTG
5151  CAGGGTCATC GCACTAAGGT CTGTGAAGAA CGGTGCTGAG ACACGGGTGG
5201  ACCTCCTCTG CACCTACCTG CAGCCCCTCA GCGGCCCAGG TCTGCCTATC
5251  AAGCAGGTGT TCCATGAGCT GAGCCAGCAG ACCCATGGCA TCACCCGGCT
5301  GGGCCCCTAC TCTCTGGACA AAGACAGCCT CTACCTTAAC GGTTACAATG
5351  AACCTGGTCT AGATGAGCCT CCTACAACTC CCAAGCCAGC CACCACATTC
5401  CTGCCTCCTC TGTCAGAAGC CACAACAGCC ATGGGGTACC ACCTGAAGAC
5451  CCTCACACTC AACTTCACCA TCTCCAATCT CCAGTATTCA CCAGATATGG
5501  GCAAGGGCTC AGCTACATTC AACTCCACCG AGGGGGTCCT TCAGCACCTG
5551  CTCAGACCCT TGTTCCAGAA GAGCAGCATG GGCCCCTTCT ACTTGGGTTG
5601  CCAACTGATC TCCCTCAGGC CTGAGAAGGA TGGGGCAGCC ACTGGTGTGG
5651  ACACCACCTG CACCTACCAC CCTGACCCTG TGGGCCCCGG GCTGGACATA
5701  CAGCAGCTTT ACTGGGAGCT GAGTCAGCTG ACCCATGGTG TCACCCAACT
5751  GGGCTTCTAT GTCCTGGACA GGGATAGCCT CTTCATCAAT GGCTATGCAC
5801  CCCAGAATTT ATCAATCCGG GGCGAGTACC AGATAAATTT CCACATTGTC
5851  AACTGGAACC TCAGTAATCC AGACCCCACA TCCTCAGAGT ACATCACCCT
5901  GCTGAGGGAC ATCCAGGACA AGGTCACCAC ACTCTACAAA GGCAGTCAAC
5951  TACATGACAC ATTCCGCTTC TGCCTGGTCA CCAACTTGAC GATGGACTCC
```

TABLE 8-continued

Complete DNA Sequence for 13 Repeats including the Carboxy Terminus of CA125 (SEQ ID NO: 49)

6001 GTGTTGGTCA CTGTCAAGGC ATTGTTCTCC TCCAATTTGG ACCCCAGCCT

6051 GGTGGAGCAA GTCTTTCTAG ATAAGACCCT GAATGCCTCA TTCCATTGGC

6101 TGGGCTCCAC CTACCAGTTG GTGGACATCC ATGTGACAGA AATGGAGTCA

6151 TCAGTTTATC AACCAACAAG CAGCTCCAGC ACCCAGCACT TCTACCCGAA

6201 TTTCACCATC ACCAACCTAC CATATTCCCA GGACAAAGCC CAGCCAGGCA

6251 CCACCAATTA CCAGAGGAAC AAAAGGAATA TTGAGGATGC GCTCAACCAA

6301 CTCTTCCGAA ACAGCAGCAT CAAGAGTTAT TTTTCTGACT GTCAAGTTTC

6351 AACATTCAGG TCTGTCCCCA ACAGGCACCA CACCGGGGTG GACTCCCTGT

6401 GTAACTTCTC GCCACTGGCT CGGAGAGTAG ACAGAGTTGC CATCTATGAG

6451 GAATTTCTGC GGATGACCCG GAATGGTACC CAGCTGCAGA ACTTCACCCT

6501 GGACAGGAGC AGTGTCCTTG TGGATGGGTA TTCTCCCAAC AGAAATGAGC

6551 CCTTAACTGG GAATTCTGAC CTTCCCTTCT GGGCTGTCAT CTTCATCGGC

6601 TTGGCAGGAC TCCTGGGACT CATCACATGC CTGATCTGCG GTGTCCTGGT

6651 GACCACCCGC CGGCGGAAGA AGGAAGGAGA ATACAACGTC CAGCAACAGT

6701 GCCCAGGCTA CTACCAGTCA CACCTAGACC TGGAGGATCT GCAA**TGA**CTG

6751 GAACTTGCCG GTGCCTGGGG TGCCTTTCCC CCAGCCAGGG TCCAAAGAAG

6801 CTTGGCTGGG GCAGAAATAA ACCATATTGG TCG

TABLE 9

Complete Amino Acid Sequence for 13 Repeats Contiguous with the Carboxy Terminus of CA125 (SEQ ID NO: 50)

1
ERVLQGLLKP LFRNSSLEYL YSGCRLASLR PEKDSSAMAV DAICTHRPDP

EDLGLDRERL YWELSNLTNG IQELGPYTLD RNSLYVNGFT HRSSMPTTST

PGTSTVDVGT SGTPSSSPSP TTAGPLLMPF TLNFTITNLQ YEEDMRRTGS

2
RKFNTMERVL QGPLSPIFKN SSVGPLYSGC RLTSLRPEKD GAATGMDAVC

LYHPNPKRPG LDREQLYWEL SQLTHNITEL GPYSLDRDSL YVNGFTHQNS

VPTTSTPGTS TVYWATTGTP SSFPGHTEPG PLLIPFTLNF TITNLQYEEN

3
MGHPGSRKFN ITERVLQGLL NPIFKNSSVG PLYSGCRLTS LRPEKDGAAT

GMDAVCLYHP NPKRPGLDRE QLYCELSQLT HNITELGPYS LDRDSLYVNG

FTHQNSVPTT STPGTSTVYW ATTGTPSSFP GHTEPGPLLI PFTLNFTITN

4
LQYEEDMRRT GSRKFNTMER VLQGLLKPLF KSTSVGPLYS GCRLTLLRPE

KHGAATGVDA ICTLRLDPTG PGLDRERLYW ELSQLTNSVT ELGPYTLDRD

SLYVNGFTHR SSVPTTSIPG TSAVHLETSG TPASLPGHTA PGPLLVPFTL

NFTITNLQYE EDMRHPGSRK FNTTERVLQG LLKPLFKSTS VGPLYSGCRL

5
TLLRPEKRGA ATGVDTICTH RLDPLNPGLD REQLYWELSK LTRGIIELGP

YLLDRGSLYV NGFTHRNFVP ITSTPGTSTV HLGTSETPSS LPRPIVPGPL

LIPFTLNFTI TNLQYEENMG HPGSRKFNIT ERVLQGLLKP LFRNSSLEYL

6
YSGCRLASLR PEKDSSAMAV DAICTHRPDP EDLGLDRERL YWELSNLTNG

IQELGPYTLD RNSLYVNGFT HRSSMPTTST PGTSTVDVGT SGTPSSSPSP

TTAGPLLMPF TLNFTITNLQ YEEDMRRTGS RKFNTMESVL QGLLKPLFKN

7
TSVGPLYSGC RLTLLRPKKD GAATGVDAIC THRLDPKSPG LNREQLYWEL

SKLTNDIEEV GPYTLDRNSL YVNGFTHRSF VAPTSTLGTS TVDLGTSGTP

SSLPSPTTGV PLLIPFTLNF TITNLQYEEN MGHPGSRKFN IMERVLQGLL

8
SPIFKNSSVG SLYSGCRLTL LRPEKDGAAT RVDAVCTHRP DPKSPGLDRE

RLYWKLSQLT HGIIELGPYT LDRHSFYVNG FTHQSSMTTT RTPDTSTMHL

ATSRTPASLS GPTTASPLLV LFTINFTITN QRYEENMHHP GSRKFNTTER

TABLE 9-continued

Complete Amino Acid Sequence for 13 Repeats Contiguous with the Carboxy Terminus of CA125 (SEQ ID NO: 50)

9

VLQGLLRPVF KNTSVGPLYS GCRLTLLRPK KDGAATKVDA ICTYRPDPKS

PGLDREQLYW ELSQLTHSIT ELGPYTQDRD SLYVNGFTHR SSVPTTSIPG

TSAVHLETSG TPASLPGPSA ASPLLVLFTL NFTITNLRYE ENMQHPGSRK

10

FNTTERVLQG LLRSLFKSTS VGPLYSGCRL TLLRPEKDGT ATGVDAICTH

HPDPKSPRLD REQLYWELSQ LTHNITELGH YALDNDSLFV NGFTHRSSVS

TTSTPGTPTV YLGASKTPAS IFGPSAASHL LILFTLNFTI TNLRYEENMW

11

PGSRKFNTTE RVLQGLLRPL FKNTSVGPLY SGSRLTLLRP EKDGEATGVD

AICTHRPDPT GPGLDREQLY LELSQLTHSI TELGPYTLDR DSLYVNGFTH

RSSVPTTSTG VVSEEPFTLN FTINNLRYMA DMGQPGSLKF NITDNVMKHL

12

LSPLFQRSSL GARYTGCRVI ALRSVKNGAE TRVDLLCTYL QPLSGPGLPI

KQVFHELSQQ THGITRLGPY SLDKDSLYLN GYNEPGLDEP PTTPKPATTF

LPPLSEATTA MGYHLKTLTL NFTISNLQYS PDMGKGSATF NSTEGVLQHL

13

LRPLFQKSSM GPFYLGCQLI SLRPEKDGAA TGVDTTCTYH PDPVGPGLDI

QQLYWELSQL THGVTQLGFY VLDRDSLFIN GYAPQNLSIR GEYQINFHIV

NWNLSNPDPT SSEYITLLRD IQDKVTTLYK GSQLHDTFRF CLVTNLTMDS

VLVTVKALFS SNLDPSLVEQ VFLDKTLNAS FHWLGSTYQL VDIHVTEMES

SVYQPTSSSS TQHFYLNFTI TNLPYSQDKA QPGTTNYQRN KRNIEDALNQ

LFRNSSIKSY FSDCQVSTFR SVPNRHHTGV DSLCNFSPLA RRVDRVAIYE

EFLRMTRNGT QLQNFTLDRS SVLVDGYSPN RNEPLTGNSD LPFWAVILIG

LAGLLGLITC LICGVLVTTR RRKKEGEYNV QQQCPGYYQS HLDLEDLQ

TABLE 10A

5' Primer Sequence for End of the Open Reading Frame for Contig #32 of Chromosome 19 Cosmid AC008734 (SEQ ID NO: 51), Primer Sequence from within the Repeat Region (SEQ ID NO: 52, 3 Primer Sets Synthesized to Piece Together Entire Open Reading Frame in Contig #32 (SEQ ID NOS: 53 thru 58), Primers to Cosmid No. AC008734 for Contig #32 (SEQ ID NOS: 59 and 60), Sense Primer Sequence (supplied by Ambion) (SEQ ID NO: 61), Anti-Sense Primer Sequence for CA125 (SEQ ID NO: 62), and 5'Sense Primer Sequence (from Ambion) (SEQ ID NO: 63) and Anti-Sense Primer Specific to CA125 (SEQ ID NO: 64)

(SEQ ID NO: 51) (5'-CAGCAGAGACCAGCACGAGTACTC-3')

(SEQ ID NO: 52) (5'-TCCACTGCCATGGCTGAGCT-3')

Primer Sets

(SEQ ID NO: 53) (Set 1) 5'-CCAGCACAGCTCTTCCCAGGAC-3'
(SEQ ID NO: 54) 5'-GGAATGGCTGAGCTGACGTCTG-3')

(SEQ ID NO: 55) (Set 2) 5'-CTTCCCAGGACAACCTCAAGG-3'
(SEQ ID NO: 56 5'-GCAGGATGAGTGAGCCACGTG-3'

(SEQ ID NO: 57) (Set 3) 5'-GTCAGATCTGGTGACCTCACTG-3'
(SEQ ID NO: 58) 5'-GAGGCACTGGAAAGCCCAGAG-3'

(SEQ ID NO: 59) 5'-CTGATGGCATTATGGAACACATCAC-3'
(SEQ ID NO: 60) 5'-CCCAGAACGAGAGACCAGTGAG-3'

(SEQ ID NO: 61) 5'-GCTGATGGCGATGAATGAACACTG-3'

(SEQ ID NO: 62) 5'-CCCAGAACGAGAGACCAGTGAG-3'

(SEQ ID NO: 63) 5'-CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG-3'
(SEQ ID NO: 64) 5'-CCTCTGTGTGCTGCTTCATTGGG-3'

TABLE 10B

Sense and Anti-Sense Primers Used to Order the CA125 Carboxy Terminal Domain (SEQ. ID NO: 303 and SEQ ID NO: 304, respectively)

(SEQ ID NO: 303) 5'-GGACAAGGTCACCACACTCTAC-3'
(SEQ ID NO: 304) 5'-GCAGATCCTCCAGGTCTAGGTGTG-3'

TABLE 10C

Sense and Anti-Sense Primers Used to Amplify Overlapping Sequences in the Repeat Domain (SEQ ID NO: 305 and SEQ ID NO: 306, respectively)

(SEQ ID NO: 305) 5' GTC TCT ATG TCA ATG GTT TCA CCC-3'
(SEQ ID NO: 306) 5'-TAG CTG CTC TCT GTC CAG TCC-3'

TABLE 11

5' Sense Primer 1 Sequence and 3' Antisense Primer 2 (SEQ ID NO: 65 and SEQ ID NO: 66, respectively), and Nucleotide and Amino Acid Sequences of the CA125 Repeat Expressed in *E. coli* (SEQ ID NO: 67 and SEQ ID NO: 68, respectively)

(SEQ ID NO: 65) 5'-ACCGGATCCATGGGCCACACAGAGCCTGGCCC-3'

(SEQ ID NO: 66) 5'-TGTAAGCTTAGGCAGGGAGGATGGAGTCC-3'

(SEQ ID NO: 67)

```
   1  ATGAGAGGAT CGCATCACCA TCACCATCAC GGATCCATGG GCCACACAGA
                                                        ↑
  51  GCCTGGCCCT CTCCTGATAC CATTCACTTT CAACTTTACC ATCACCAACC
 101  TGCATTATGA GGAAAACATG CAACACCCTG GTTCCAGGAA GTTCAACACC
 151  ACGGAGAGGG TTCTGCAGGG TCTGCTCAAG CCCTTGTTCA AGAACACCAG
 201  TGTTGGCCCT CTGTACTCTG GCTGCAGACT GACCTTGCTC AGACCTGAGA
 251  AGCATGAGGC AGCCACTGGA GTGGACACCA TCTGTACCCA CCGCGTTGAT
 301  CCCATCGGAC CTGGACTGGA CAGAGAGCGG CTATACTGGG AGCTGAGCCA
 351  GCTGACCAAC AGCATCACAG AGCTGGGACC CTACACCCTG GACAGGGACA
 401  GTCTCTATGT CAATGGCTTC AACCCTCGGA GCTCTGTGCC AACCACCAGC
 451  ACTCCTGGGA CCTCCACAGT GCACCTGGCA ACCTCTGGGA CTCCATCCTC
 501  CCTGCCT
```

(SEQ ID NO: 68)

```
M R G S H H H H H H G S M G H T E P G P L L I P F T F N F T I T N L
H Y E E N M Q H P G S R K F N T T E R V L Q G L L K P L F K N T S V
G P L Y S G C R L T L L R P E K H E A A T G V D T I C T H R V D P I
G P G L D R E R L Y W E L S Q L T N S I T E L G P Y T L D R D S L Y
V N G F N P R S S V P T T S T P G T S T V H L A T S G T P S S L P
```

TABLE 12

Additional Multiple Repeat Amino Acid Sequences
(SEQ ID NO: 69 thru SEQ ID NO: 80)

(SEQ ID NO: 69)

ERVLQGLLGP MFKNTSVGLL YSGCRLTLLR PKKDGAATKV DAICTYRPDP
KSPGLDREQL YWELSQLTHS ITELGPYTLD RDSLYVNGFT QRSSVPTTSI
PGTPTVDLGT SGTPVSKPGP SAASPLLIPF TINFTITNLR YEENMGHPGS
RKFNIMERVL QGLLKPLFKN TSVGPLYSGC RLTLLRPKKD GAATGVDAIC
THRLDPKSPG LNREQLYWEL SKLTNDIEEL GPYTLDRNSL YVNGFTHQSS
VSTTSTPGTS TVDLRTSGTP SSLSSPTIMA AGPLLIPFTI NFTITNLRYE
ENMHHPGSRK FNTMERVLQG LLMPLFKNTS VSSLYSGCRL TLLRPEKDGA
ATRVDAVCTH RPDPKSPGLD RERLYWKLSQ LTHGITELGP YTLDRNSLYV
NGFTHRSSMP TTSTPGTSTV DVGTSGTPSS SPSPTTAGPL LMPFTLNFTI
TNLQYEEDMR RTGSRKFNTM ERVLQGLLKP LFKSTSVGPL YSGCRLTLLR
PEKHGAATGV DAICTLRLDP TGPGLDRERL YWELSQLTNS VTELGPYTLD
RDSLYVNGFT HRSSVPTTSI PGTSAVHLET SGTPASLPGH TAPGPLLIPF
TLNFTITNLH YEENMQHPGS RKFNTMERVL QGCLVPCSRN TNVGLLYSGC
RLTLLRXEKX XAATXVDXXC XXXXDPXXPG LDREXLYWEL SXLTXXIXEL
GPYTLDRNSL YVNGFTHRSS VAPTSTPGTS TVDLGTSGTP SSLPSPTTVP
LLVPFTLNFT ITNLQYGEDM RHPGSRKFNT TERVLQGLLG PLFKNSSVGP
LYSGCRLISL RSEKDGAATG VDAICTHHLN PQSPGLDREQ LYWQLSQVTN
GIKELGPYTL DRNSLYVNGF THRSSGLTTS TPWTSTVDLG TSGTPSPVPS
PTTAGPLLI

TABLE 12-continued

Additional Multiple Repeat Amino Acid Sequences
(SEQ ID NO: 69 through SEQ ID NO: 80)

(SEQ ID NO: 70)

QGLLGPMFKN TSVGLLYSGC RLTLLRPEKR GAATGVDTIC THRLDPLNPG

LDREQLYWEL SKLTRGIIEL GPYLLDRGSL YVNGFTHRNF VPITSTPGTS

TVHLGTSETP SSLPRPIVPG PLLVPFTLNF TITNLQYEEA MRHPGSRKFN

TTERVLQGLL RPLFKNTSVS SLYSGCRLTL LRPEKDGAAT RVDAACTYRP

DPKSPGLDRE QLYWELSQLT HSITELGPYT LDRVSLYVNG FNPRSSVPTT

STPGTSTVHL ATSGTPSSLP GHTAPVPLLI PFTLNFTITN LQYEEDMRHP

GSRKFNTMER VLQGLLRPLF KNTSIGPLYS SCRLTLLRPE KDKAATRVDA

ICTHHPDPQS PGLNREQLYW ELSQLTHGIT ELGPYTLDRD SLYVDGFTHW

SPIPTTSTPG TSIVNLGTSG IPPSLPETTA TGPLLIPFTP NFTITNLQYE

EDMRRTGSRK FNTMERVLQG LLSPIFKNSS VGPLYSGCRL TSLRPEKDGA

ATGMDAVCLY HPNPKRPGLD REQLY

(SEQ ID NO:71)

ERVLQGLLKP LFKSTSVGPL YSGCRLTLLR PEKDGVATRV DAICTHRPDP

KIPGLDRQQL YWELSQLTHS ITELGPYTLD RDSLYVNGFT QRSSVPTTST

PGTFTVQPET SETPSSLPGP TATGPVLLPF TLNFTIINLQ YEEDMHRPGS

RKFNTTERVL QGLLMPLFKN TSVGPLYSGC RLTLLRPEKQ EAATGVDTIC

THRLDPSEPG LDREQLYWEL SQLTNSITEL GPYTLDRDSL YVNGFTHSGV

LCPPPSILGI FTVQPETFET PSSLPGPTAT GPVLLPFTLN FTIINLQYEE

DMHRPGSRKF NTTERVLQGL LTPLFKNTSV GPLYSGCRLT LLRPEKQEAA

TGVDTICTHR VDPIGPGLDR ERLYWELSQL TNSITELGPY TLDRDSLYVN

GFNPWSSVPT TSTPGTSTVH LATSGTPSSL PGHTAPVPLL IPFTLNFTIT

TABLE 12-continued

Additional Multiple Repeat Amino Acid Sequences (SEQ ID NO: 69 through SEQ ID NO: 80)

NLHYEENMQH PGSRKFNTTE RVLQGLLKPL FKSTSVGPLY SGCRLTLLRP

EKHGAATGVD AICTHRLDPK SPGVDREQLY WELSQLTNGI KELGPYTLDR

NSLYVNGFTH WIPVPTSSTP GTSTVDLGSG TPSSLPSPTT AGPL

(SEQ ID NO: 72)

TSVGPLYSGC RLTLLRSEKD GAATGVDAIY THRLDPKSPG VDREQLYWEL

SQLTNGIKEL GPYTLDRNSL YVNGFTHQTS APNTSTPGTS TVDLGTSGTP

SSLPSPTSAG PLLIPFTINF TITNLRYEEN MHHPGSRKFN TMERVLQGLL

KPLFKSTSVG PLYSGCRLTL LRPEKDGVAT RVDAICTHRP DPKIPGLDRQ

QLYWELSQLT HSITELGPYT LDRDSLYVNG FTQRSSVPTT STPGTFTVQP

ETSETPSSLP GPTATGPVLL PFTLNFTIIN LQYEEDMHRP GSRKFNTTER

VLQGLLKPLF KSTSVGPLYS GCRLTLLRPE KHGAATGVDA ICTLRLDPTG

PGLDRERLYW ELSQLTNSIT ELGPYTLDRD SLYVNGFNPW SSVPTTSTPG

TSTVHLATSG TPSSLPGHTA PVPL

(SEQ ID NO:73)

ERVLQGLLKP LFKSTSVGPL YSGCRLTLLR PEKRGAATGV DTICTHRLDP

LNPGLDREQL YWELSKLTRG IIELGPYLLD RDSLYVNGFT HRSSVPTTSI

PGTSAVHLET SGTPASLPGH TAPGPLLVPF TLNFTITNLQ YEEDMRHPGS

RKFNTTERVL QGLLKPLFKS TSVGPLYSGC RLTLLRPEKR GAATGVDTIC

THRLDPLNPG LDREQLYWEL SKLTRGIIEL GPYLLDRGSL YVNGFTHRNF

VPITSTPGTS TVHLGTSETP SSLPRPIVPG PLLIPF

TABLE 12-continued

Additional Multiple Repeat Amino Acid Sequences
(SEQ ID NO: 69 through SEQ ID NO: 80)

(SEQ ID NO: 74)

ERVLQGLLRP VFKNTSVGPL YSGCRLTLLR PKKDGAATKV DAICTYRPDP

KSPGLDREQL YWELSQLTHS ITELGPYTLD RDSLYVNGFT QRSSVPTTSI

PGTPTVDLGT SGTPVSKPGP SAASPLLVPF TLNFTITNLQ YEEDMHRPGS

RKFNATERVL QGLLSPIFKN SSVGPLYSGC RLTSLRPEKD GAATGMDAVC

LYHPNPKRPG LDREQLYWEL SQLTHNITEL GPYSLDRDSL YVNGFTHQSS

MTTTRTPDTS TMHLATSRTP ASLSGPTTAS PLLIPF

(SEQ ID NO: 75)

ERVLQGLLKP LFKSTSVGPL YSGCRLTLLR PEKRGAATGV DTICTHRLDP

LNPGLDREQL YWELSKLTRG IIELGPYLLD RGSLYVNGFS RQSSMTTTRT

PDTSTMHLAT SRTPASLSGP TTASPLLIPF TLNFTITNLQ YEENMGHPGS

RKFNIMERVL QGLLNPIFKN SSVGPLYSGC RLTSLKPEKD GAATGMDAVC

LYHPNPKRPG LDREQLYWEL SQLTHGIKEL GPYTLDRNSL YVNGFTHRSS

VAPTSTPGTS TVDLGTSGTP SSLPSPTTAV PLLIPF

(SEQ ID NO: 76)

ERVLQGLLKP LFRNSSLEYL YSGCRLASLR PEKDSSAMAV DAICTHRPDP

EDLGLDRERL YWELSNLTNG IQELGPYTLD RNSLYVNGFT HRSSGLTTST

PWTSTVDLGT SGTPSPVPSP TTAGPLLIPF TLNFTITNLQ YEENMGHPGS

RKFNIMERVL QGLLMPLFKN TSVSSLYSGC RLTLLRPEKD GAATRVDAVC

TQRPDPKSPG LDRERLYWKL SQLTHGITEL GPYTLDRHSL YVNGLTHQSS

MTTTRTPDTS TMHLATSRTP ASLSGPTTAS PLLIPF

TABLE 12-continued

Additional Multiple Repeat Amino Acid Sequences
(SEQ ID NO: 69 through SEQ ID NO: 80)

(SEQ ID NO: 77)

ERVLQGLLSP ISKNSSVGPL YSGCRLTSLR PEKDGAATGM DAVCLYHPNP

KRPGLDREQL YWELSQLTHN ITELGPYSLD RDSLYVNGFT HQNSVPTTST

PGTSTVYWAT TGTPSSFPGH TEPGPLLIPF TVNFTITNLR YEENMHHPGS

RKFNTTERVL QGLLRPVFKN TSVGPLYSGC RLTLLRPKKD GAATKVDAIC

TYRPDPKSPG LDREQLYWEL SKLTNDIEEL GPYTLDRNSL YVNGFTHQSS

VSTTSTPGTS TVDLRTSGTP SSLSSPTIMA AGPLLIPF

(SEQ ID NO: 78)

ERVLHGLLTP LFKNTRVGPL YSGCRLTLLR PEKQEAATGV DTICTHRVDP

IGPGLDRERL YWELSQLTNS ITELGPYTLD RDSLYVNGFN PWSSVPTTST

PGTSTVHLAT SGTPSSLPGH TAPVPLLIPF TLNFTITNLH YEENMQHPGS

RKFNTTERVL QGLLKPLFKN TSVGPLYSGC RLTLFKPEKH EAATGVDAIC

TLRLDPTGPG LDRQLYWELS QLTNSVTELG PYTLDRDSLY VNGFTHRSSV

PTTSIPGTSA VHLETSGTPA SLPGHTAPGP LLIPFTLNFT ITNLQYEEDM

RRTGSRKFNT MERVLQGLLK PLFKSTSVGP LYSGCRLTLL RPEKRGAATG

VDTICTHRLD PLNPGLDREQ LYWELSKLTR GIIELGPYLL DRGSLYVNGF

THRNFVPITS TPGTSTVHLG TSETPSSLPR PIVPGPLLIP FTINFTITNL

RYEENMHHPG SRKFNIMERV LQGLLGPLFK NSSVGPLYSG CRLISLRSEK

DGAATGVDAI CTHHLNPQSP GLDREQLYWQ LSQMTNGIKE LGPYTLDRNS

LYVNGFTHRS SGLTTSTPWT STVDLGTSGT PSPVPSPTTA GPLLIPF

TABLE 12-continued

Additional Multiple Repeat Amino Acid Sequences (SEQ ID NO: 69 through SEQ ID NO: 80)

(SEQ ID NO: 79)

GPLYSGCRLT SLRPEKDGAA TGMDAVCLYH PNPKRPGLDR EQLYWELSQL

THNITELGPY SLDRDSLYVN GFTHQNSVPT TSTPGTSTVY WATTGTPSSF

PGHTEPGPLL IPFTLNFTIT NLQYEENMGH PGSRKFNITE SVLQGLLTPL

FKNSSVGPLY SGCRLISLRS EKDGAATGVD AICTHHLNPQ SPGLDREQLY

WQLSQMTNGI KELGPYTLDR DSLYVNGFTH RSLGLTTSTP WTSTVDLGTS

GTPSPVPSPT TAGPLLIPFT LNFTITNLQY EENMGHPGSR KFNIMERVLQ

GLLRPVFKNT SVGPLYSGCR LTLLRPKKDG AATKVDAICT YRPDPKSPGL

DREQLYWELS QLTHSITELG PYTLDRDSLY VNGFTQRSSV PTTSIPGTPT

VDLGTSGTPV SKPGPSAASP

(SEQ ID NO: 80)

QLYWELSKLT NDIEELGPYT LDRNSLYVNG FTHQSSVSTT STPGTSTVDL

RTSGTPSSLS SPTIMAAGPL LIPFTLNFTI TNLQYEENMG HPGSRKFNIM

ERVLQGLLGP MFKNTSVGLL YSGCRLTLLR PEKNGAATGM DAICSHRLDP

KSPGLNREQL YWELSQLTHG IKELGPYTLD RNSLYVNGFT HRSSVAPTST

PGTSTVDLGT SGTPSSLPSP TTAVPLLIPF TLNFTITNLK YEEDMHCPGS

RKFNTTERVL QSLFGPMFKN TSVGPLYSGC RLTLLRSEKD GAATGVDAIC

THRLDPKSLG VDREQLYWEL SQLTNGIKEL GPYTLDRNSL YVNGFTHQTS

APNTSTPGTS TVDLGTSGTP SSLPSPTSAG PLLVPFTLNF TITNLQYEED

MRRTGSRKFN TMESVLQGLL KPLFKNTSVG PLYSGCRLTL LRPEKDGAAT

GVDAICTHRL DPKSPGLNRE QLYWELSKL

TABLE 13

Amino Terminal Nucleotide Sequence
(SEQ ID NO: 81)

```
   1  CAGAGAGCGT TGAGCTGGGA ACAGTGACAA GTGCTTATCA AGTTCCTTCA
  51  CTCTCAACAC GGTTGACAAG AACTGATGGC ATTATGGAAC ACATCACAAA
 101  AATACCCAAT GAAGCAGCAC ACAGAGGTAC CATAAGACCA GTCAAAGGCC
 151  CTCAGACATC CACTTCGCCT GCCAGTCCTA AAGGACTACA CACAGGAGGG
 201  ACAAAAAGAA TGGAGACCAC CACCACAGCT TTGAAGACCA CCACCACAGC
 251  TTTGAAGACC ACTTCCAGAG CCACCTTGAC CACCAGTGTC TATACTCCCA
 301  CTTTGGGAAC ACTGACTCCC CTCAATGCAT CAAGGCAAAT GGCCAGCACA
 351  ATCCTCACAG AAATGATGAT CACAACCCCA TATGTTTTCC CTGATGTTCC
 401  AGAAACGACA TCCTCATTGG CTACCAGCCT GGGAGCAGAA ACCAGCACAG
 451  CTCTTCCCAG GACAACCCCA TCTGTTCTCA ATAGAGAATC AGAGACCACA
 501  GCCTCACTGG TCTCTCGTTC TGGGGCAGAG AGAAGTCCGG TTATTCAAAC
 551  TCTAGATGTT TCTTCTAGTG AGCCAGATAC AACAGCTTCA TGGGTTATCC
 601  ATCCTGCAGA GACCATCCCA ACTGTTTCCA AGACAACCCC CAATTTTTTC
 651  CACAGTGAAT TAGACACTGT ATCTTCCACA GCCACCAGTC ATGGGGCAGA
 701  CGTCAGCTCA GCCATTCCAA CAAATATCTC ACCTAGTGAA CTAGATGCAC
 751  TGACCCCACT GGTCACTATT TCGGGGACAG ATACTAGTAC AACATTCCCA
 801  ACACTGACTA AGTCCCCACA TGAAACAGAG ACAAGAACCA CATGGCTCAC
 851  TCATCCTGCA GAGACCAGCT CAACTATTCC CAGAACAATC CCCAATTTTT
 901  CTCATCATGA ATCAGATGCC ACACCTTCAA TAGCCACCAG TCCTGGGGCA
 951  GAAACCAGTT CAGCTATTCC AATTATGACT GTCTCACCTG GTGCAGAAGA
```

TABLE 13-continued

Amino Terminal Nucleotide Sequence (SEQ ID NO: 81)	
1001	TCTGGTGACC TCACAGGTCA CTAGTTCTGG GACAGACAGA AATATGACTA
1051	TTCCAACTTT GACTCTTTCT CCTGGTGAAC CAAAGACGAT AGCCTCATTA
1101	GTCACCCATC CTGAAGCACA GACAAGTTCG GCCATTCCAA CTTCAACTAT
1151	CTCGCCTGCT GTATCACGGT TGGTGACCTC AATGGTCACC AGTTTGGCGG
1201	CAAAGACAAG TACAACTAAT CGAGCTCTGA CAAACTCCCC TGGTGAACCA
1251	GCTACAACAG TTTCATTGGT CACGCATCCT GCACAGACCA GCCCAACAGT
1301	TCCCTGGACA ACTTCCATTT TTTTCCATAG TAAATCAGAC ACCACACCTT
1351	CAATGACCAC CAGTCATGGG GCAGAATCCA GTTCAGCTGT TCCAACTCCA
1401	ACTGTTTCAA CTGAGGTACC AGGAGTAGTG ACCCCTTTGG TCACCAGTTC
1451	TAGGGCAGTG ATCAGTACAA CTATTCCAAT TCTGACTCTT TCTCCTGGTG
1501	AACCAGAGAC CACACCTTCA ATGGCCACCA GTCATGGGGA AGAAGCCAGT
1551	TCTGCTATTC CAACTCCAAC TGTTTCACCT GGGGTACCAG GAGTGGTGAC
1601	CTCTCTGGTC ACTAGTTCTA GGGCAGTGAC TAGTACAACT ATTCCAATTC
1651	TGACTTTTTC TCTTGGTGAA CCAGAGACCA CACCTTCAAT GGCCACCAGT
1701	CATGGGACAG AAGCTGGCTC AGCTGTTCCA ACTGTTTTAC CTGAGGTACC
1751	AGGAATGGTG ACCTCTCTGG TTGCTAGTTC TAGGGCAGTA ACCAGTACAA
1801	CTCTTCCAAC TCTGACTCTT TCTCCTGGTG AACCAGAGAC CACACCTTCA
1851	ATGGCCACCA GTCATGGGGC AGAAGCCAGC TCAACTGTTC CAACTGTTTC
1901	ACCTGAGGTA CCAGGAGTGG TGACCTCTCT GGTCACTAGT TCTAGTGGAG
1951	TAAACAGTAC AAGTATTCCA ACTCTGATTC TTTCTCCTGG TGAACTAGAA

TABLE 13-continued

Amino Terminal Nucleotide Sequence (SEQ ID NO: 81)	
2001	ACCACACCTT CAATGGCCAC CAGTCATGGG GCAGAAGCCA GCTCAGCTGT
2051	TCCAACTCCA ACTGTTTCAC CTGGGGTATC AGGAGTGGTG ACCCCTCTGG
2101	TCACTAGTTC CAGGGCAGTG ACCAGTACAA CTATTCCAAT TCTAACTCTT
2151	TCTTCTAGTG AGCCAGAGAC CACACCTTCA ATGGCCACCA GTCATGGGGT
2201	AGAAGCCAGC TCAGCTGTTC TAACTGTTTC ACCTGAGGTA CCAGGAATGG
2251	TGACCTCTCT GGTCACTAGT TCTAGAGCAG TAACCAGTAC AACTATTCCA
2301	ACTCTGACTA TTTCTTCTGA TGAACCAGAG ACCACAACTT CATTGGTCAC
2351	CCATTCTGAG GCAAAGATGA TTTCAGCCAT TCCAACTTTA GCTGTCTCCC
2401	CTACTGTACA AGGGCTGGTG ACTTCACTGG TCACTAGTTC TGGGTCAGAG
2451	ACCAGTGCGT TTTCAAATCT AACTGTTGCC TCAAGTCAAC CAGAGACCAT
2501	AGACTCATGG GTCGCTCATC CTGGGACAGA AGCAAGTTCT GTTGTTCCAA
2551	CTTTGACTGT CTCCACTGGT GAGCCGTTTA CAAATATCTC ATTGGTCACC
2601	CATCCTGCAG AGAGTAGCTC AACTCTTCCC AGGACAACCT CAAGGTTTTC
2651	CCACAGTGAA TTAGACACTA TGCCTTCTAC AGTCACCAGT CCTGAGGCAG
2701	AATCCAGCTC AGCCATTTCA ACTACTATTT CACCTGGTAT ACCAGGTGTG
2751	CTGACATCAC TGGTCACTAG CTCTGGGAGA GACATCAGTG CAACTTTTCC
2801	AACAGTGCCT GAGTCCCCAC ATGAATCAGA GGCAACAGCC TCATGGGTTA
2851	CTCATCCTGC AGTCACCAGC ACAACAGTTC CCAGGACAAC CCCTAATTAT
2901	TCTCATAGTG AACCAGACAC CACACCATCA ATAGCCACCA GTCCTGGGGC
2951	AGAAGCCACT TCAGATTTTC CAACAATAAC TGTCTCACCT GATGTACCAG

TABLE 13-continued

Amino Terminal Nucleotide Sequence
(SEQ ID NO: 81)

```
3001  ATATGGTAAC CTCACAGGTC ACTAGTTCTG GGACAGACAC CAGTATAACT
3051  ATTCCAACTC TGACTCTTTC TTCTGGTGAG CCAGAGACCA CAACCTCATT
3101  TATCACCTAT TCTGAGACAC ACACAAGTTC AGCCATTCCA ACTCTCCCTG
3151  TCTCCCCTGG TGCATCAAAG ATGCTGACCT CACTGGTCAT CAGTTCTGGG
3201  ACAGACAGCA CTACAACTTT CCCAACACTG ACGGAGACCC CATATGAACC
3251  AGAGACAACA GCCATACAGC TCATTCATCC TGCAGAGACC AACACAATGG
3301  TTCCCAAGAC AACTCCCAAG TTTTCCCATA GTAAGTCAGA CACCACACTC
3351  CCAGTAGCCA TCACCAGTCC TGGGCCAGAA GCCAGTTCAG CTGTTTCAAC
3401  GACAACTATC TCACCTGATA TGTCAGATCT GGTGACCTCA CTGGTCCCTA
3451  GTTCTGGGAC AGACACCAGT ACAACCTTCC CAACATTGAG TGAGACCCCA
3501  TATGAACCAG AGACTACAGT CACGTGGCTC ACTCATCCTG CAGAAACCAG
3551  CACAACGGTT TCTGGGACAA TTCCCAACTT TTCCCATAGG GGATCAGACA
3601  CTGCACCCTC AATGGTCACC AGTCCTGGAG TAGACACGAG GTCAGGTGTT
3651  CCAACTACAA CCATCCCACC CAGTATACCA GGGGTAGTGA CCTCACAGGT
3701  CACTAGTTCT GCAACAGACA CTAGTACAGC TATTCCAACT TTGACTCCTT
3751  CTCCTGGTGA ACCAGAGACC ACAGCCTCAT CAGCTACCCA TCCTGGGACA
3801  CAGACTGGCT TCACTGTTCC AATTCGGACT GTTCCCTCTA GTGAGCCAGA
3851  TACAATGGCT TCCTGGGTCA CTCATCCTCC ACAGACCAGC ACACCTGTTT
3901  CCAGAACAAC CTCCAGTTTT TCCCATAGTA GTCCAGATGC CACACCTGTA
3951  ATGGCCACCA GTCCTAGGAC AGAAGCCAGT TCAGCTGTAC TGACAACAAT
```

TABLE 13-continued

Amino Terminal Nucleotide Sequence (SEQ ID NO: 81)	
4001	CTCACCTGGT GCACCAGAGA TGGTGACTTC ACAGATCACT AGTTCTGGGG
4051	CAGCAACCAG TACAACTGTT CCAACTTTGA CTCATTCTCC TGGTATGCCA
4101	GAGACCACAG CCTTATTGAG CACCCATCCC AGAACAGGGA CAAGTAAAAC
4151	ATTTCCTGCT TCAACTGTGT TTCCTCAAGT ATCAGAGACC ACAGCCTCAC
4201	TCACCATTAG ACCTGGTGCA GAGACTAGCA CAGCTCTCCC AACTCAGACA
4251	ACATCCTCTC TCTTCACCCT ACTTGTAACT GGAACCAGCA GAGTTGATCT
4301	AAGTCCAACT GCTTCACCTG GTGTTTCTGC AAAAACAGCC CCACTTTCCA
4351	CCCATCCAGG GACAGAGACC AGCACAATGA TTCCAACTTC AACTCTTTCC
4401	CTTGGTTTAC TAGAGACTAC AGGCTTACTG GCCACCAGCT CTTCAGCAGA
4451	GACCAGCACG AGTACTCTAA CTCTGACTGT TTCCCCTGCT GTCTCTGGGC
4501	TTTCCAGTGC CTCTATAACA ACTGATAAGC CCCAAACTGT GACCTCCTGG
4551	AACACAGAAA CCTCACCATC TGTAACTTCA GTTGGACCCC CAGAATTTTC
4601	CAGGACTGTC ACAGGCACCA CTATGACCTT GATACCATCA GAGATGCCAA
4651	CACCACCTAA AACCAGTCAT GGAGAAGGAG TGAGTCCAAC CACTATCTTG
4701	AGAACTACAA TGGTTGAAGC CACTAATTTA GCTACCACAG GTTCCAGTCC
4751	CACTGTGGCC AAGACAACAA CCACCTTCAA TACACTGGCT GGAAGCCTCT
4801	TTACTCCTCT GACCACACCT GGGATGTCCA CCTTGGCCTC TGAGAGTGTG
4851	ACCTCAAGAA CAAGTTATAA CCATCGGTCC TGGATCTCCA CCACCAGCAG
4901	TTATAACCGT CGGTACTGGA CCCCTGCCAC CAGCACTCCA GTGACTTCTA
4951	CATTCTCCCC AGGGATTTCC ACATCCTCCA TCCCCAGCTC CACAGCAGCC

TABLE 13-continued

Amino Terminal Nucleotide Sequence (SEQ ID NO: 81)	
5001	ACAGTCCCAT TCATGGTGCC ATTCACCCTC AACTTCACCA TCACCAACCT
5051	GCAGTACGAG GAGGACATGC GGCACCCTGG TTCCAGGAAG TTCAACGCCA
5101	CAGAGAGAGA ACTGCAGGGT CTGCTCAAAC CCTTGTTCAG GAATAGCAGT
5151	CTGGAATACC TCTATTCAGG CTGCAGACTA GCCTCACTCA GGCCAGAGAA
5201	GGATAGCTCA GCCATGGCAG TGGATGCCAT CTGCACACAT CGCCCTGACC
5251	CTGAAGACCT CGGACTGGAC AGAGAGCGAC TGTACTGGGA GCTGAGCAAT
5301	CTGACAAATG GCATCCAGGA GCTGGGCCCC TACACCCTGG ACCGGAACAG
5351	TCTCTATGTC AATGGTTTCA CCCATCGAAG CTCTATGCCC ACCACCAGCA
5401	CTCCTGGGAC CTCCACAGTG GATGTGGGAA CCTCAGGGAC TCCATCCTCC
5451	AGCCCCAGCC CCACG

TABLE 14

Amino Terminal Protein Sequence
(SEQ ID NO: 82)

```
  1  ESVLEGTVTS AYQVPSLSTR LTRTDGIMEH ITKIPNEAAH RGTIRPVKGP
 51  QTSTSPASPK GLHTGGTKRM ETTTTALKTT TTALKTTSRA TLTTSVYTPT
101  LGTLTPLNAS RQMASTILTE MMITTPYVFP DVPETTSSLA TSLGAETSTA
151  LPRTTPSVLN RESETTASLV SRSGAERSPV IQTLDVSSSE PDTTASWVIH
201  PAETIPTVSK TTPNFFHSEL DTVSSTATSH GADVSSAIPT NISPSELDAL
251  TPLVTISGTD TSTTFPTLTK SPHETETRTT WLTHPAETSS TIPRTIPNFS
301  HHESDATPSI ATSPGAETSS AIPIMTVSPG AEDLVTSQVT SSGTDRNMTI
351  PTLTLSPGEP KTIASLVTHP EAQTSSAIPT STISPAVSRL VTSMVTSLAA
401  KTSTTNRALT NSPGEPATTV SLVTHPAQTS PTVPWTTSIF FHSKSDTTPS
451  MTTSHGAESS SAVPTPTVST EVPGVVTPLV TSSRAVISTT IPILTLSPGE
501  PETTPSMATS HGEEASSAIP TPTVSPGVPG VVTSLVTSSR AVTSTTIPIL
551  TFSLGEPETT PSMATSHGTE AGSAVPTVLP EVPGMVTSLV ASSRAVTSTT
601  LPTLTLSPGE PETTPSMATS HGAEASSTVP TVSPEVPGVV TSLVTSSSGV
651  NSTSIPTLIL SPGELETTPS MATSHGAEAS SAVPTPTVSP GVSGVVTPLV
701  TSSRAVTSTT IPILTLSSSE PETTPSMATS HGVEASSAVL TVSPEVPGMV
751  TSLVTSSRAV TSTTIPTLTI SSDEPETTTS LVTHSEAKMI SAIPTLAVSP
801  TVQGLVTSLV TSSGSETSAF SNLTVASSQP ETIDSWVAHP GTEASSVVPT
851  LTVSTGEPFT NISLVTHPAE SSSTLPRTTS RFSHSELDTM PSTVTSPEAE
901  SSSAISTTIS PGIPGVLTSL VTSSGRDISA TFPTVPESPH ESEATASWVT
```

TABLE 14-continued

Amino Terminal Protein Sequence (SEQ ID NO: 82)
951 HPAVTSTTVP RTTPNYSHSE PDTTPSIATS PGAEATSDFP TITVSPDVPD
1001 MVTSQVTSSG TDTSITIPTL TLSSGEPETT TSFITYSETH TSSAIPTLPV
1051 SPGASKMLTS LVISSGTDST TTFPTLTETP YEPETTAIQL IHPAETNTMV
1101 PRTTPKFSHS KSDTTLPVAI TSPGPEASSA VSTTTISPDM SDLVTSLVPS
1151 SGTDTSTTFP TLSETPYEPE TTATWLTHPA ETSTTVSGTI PNFSHRGSDT
1201 APSMVTSPGV DTRSGVPTTT IPPSIPGVVT SQVTSSATDT STAIPTLTPS
1251 PGEPETTASS ATHPGTQTGF TVPIRTVPSS EPDTMASWVT HPPQTSTPVS
1301 RTTSSFSHSS PDATPVMATS PRTEASSAVL TTISPGAPEM VTSQITSSGA
1351 ATSTTVPTLT HSPGMPETTA LLSTHPRTET SKTFPASTVF PQVSETTASL
1401 TIRPGAETST ALPTQTTSSL FTLLVTGTSR VDLSPTASPG VSAKTAPLST
1451 HPGTETSTMI PTSTLSLGLL ETTGLLATSS SAETSTSTLT LTVSPAVSGL
1501 SSASITTDKP QTVTSWNTET SPSVTSVGPP EFSRTVTGTT MTLIPSEMPT
1551 PPKTSHGEGV SPTTILRTTM VEATNLATTG SSPTVAKTTT TFNTLAGSLF
1601 TPLTTPGMST LASESVTSRT SYNHRSWIST TSSYNRRYWT PATSTPVTST
*
1651 FSPGISTSSI PSSTAATVPF MVPFTLNFTI TNLQYEEDMR HPGSRKFNAT
1701 ERELQGLLKP LFRNSSLEYL YSGCRLASLR PEKDSSAMAV DAICTHRPDP
1751 EDLGLDRERL YWELSNLTNG IQELGPYTLD RNSLYVNGFT HRSSMPTTST
1801 PGTSTVDVGT SGTPSSSPSP T

TABLE 15

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

(SEQ ID NO: 83)

```
   1  GCCACAGTCC CATTCATGGT GCCATTCACC CTCAACTTCA CCATCACCAA
  51  CCTGCAGTAC GAGGAGGACA TGCGGCACCC TGGTTCCAGG AAGTTCAACG
 101  CCACAGAGAG AGAACTGCAG GGTCTGCTCA AACCCTTGTT CAGGAATAGC
 151  AGTCTGGAAT ACCTCTATTC AGGCTGCAGA CTAGCCTCAC TCAGGCCAGA
 201  GAAGGATAGC TCAGCCATGG CAGTGGATGC CATCTGCATA CATCGCCCTG
 251  ACCCTGAAGA CCTCGGACTG GACAGAGAGC GACTGTACTG GGAGCTGAGC
 301  AATCTGACAA ATGGCATCCA GGAGCTGGGC CCCTACACCC TGGACCGGAA
 351  CAGTCTCTAT GTCAATGGTT TCACCCATCG AAGCTCTATG CCCACCACCA
 401  GCACTCCTGG GACCTCCACA GTGGATGTGG GAACCTCAGG GACTCCATCC
 451  TCCAGCCCCA GCCCCACG
```

(SEQ ID NO: 84)

```
   1  GCTGCTGGCC CTCTCCTGAT GCCGTTCACC CTCAACTTCA CCATCACCAA
  51  CCTGCAGTAC GAGGAGGACA TGCGTCGCAC TGGCTCCAGG AAGTTCAACA
 101  CCATGGAGAG TGTCCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAACACC
 151  AGTGTTGGCC CTCTGTACTC TGGCTGCAGA TTGACCTTGC TCAGGCCCAA
 201  GAAAGATGGG GCAGCCACTG GAGTGGATGC CATCTGCACC CACCGCCTTG
 251  ACCCCAAAAG CCCTGGACTC AACAGGGAGC AGCTGTACTG GGAGCTAAGC
 301  AAACTGACCA ATGACATTGA AGAGCTGGGC CCCTACACCC TGGACAGGAA
 351  CAGTCTCTAT GTCAATGGTT TCACCCATCA GAGCTCTGTG TCCACCACCA
 401  GCACTCCTGG GACCTCCACA GTGGATCTCA GAACCTCAGG GACTCCATCC
```

TABLE 15- continued

CA125 Repeat Nucleotide Sequence (SEQ ID NO: 83 thru SEQ ID NO: 145)

451 TCCCTCTCCA GCCCCACAAT TATG

(SEQ ID NO: 85)

1 GCTGCTGGCC CTCTCCTGGT ACCATTCACC CTCAACTTCA CCATCACCAA

51 CCTGCAGTAT GGGGAGGACA TGGGTCACCC TGGCTCCAGG AAGTTCAACA

101 CCACAGAGAG GGTCCTGCAG GGTCTGCTTG GTCCCATATT CAAGAACACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTCTC TCAGGTCTGA

201 GAAGGATGGA GCAGCCACTG GAGTGGATGC CATCTGCATC CATCATCTTG

251 ACCCCAAAAG CCCTGGACTC AACAGAGAGC GGCTGTACTG GGAGCTGAGC

301 CAACTGACCA ATGGCATCAA AGAGCTGGGC CCCTACACCC TGGACAGGAA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCG GACCTCTGTG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGGACCTTG GAACCTCAGG GACTCCATTC

451 TCCCTCCCAA GCCCCGCA

(SEQ ID NO: 86)

1 ACTGCTGGCC CTCTCCTGGT GCTGTTCACC CTCAACTTCA CCATCACCAA

51 CCTGAAGTAT GAGGAGGACA TGCATCGCCC TGGCTCCAGG AAGTTCAACA

101 CCACTGAGAG GGTCCTGCAG ACTCTGCTTG GTCCTATGTT CAAGAACACC

151 AGTGTTGGCC TTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGTCCGA

201 GAAGGATGGA GCAGCCACTG GAGTGGATGC CATCTGCACC CACCGTCTTG

251 ACCCCAAAAG CCCTGGACTG GACAGAGAGC AGCTATACTG GGAGCTGAGC

301 CAGCTGACCA ATGGCATCAA AGAGCTGGGC CCCTACACCC TGGACAGGAA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

351 CAGTCTCTAT GTCAATGGTT TCACCCATTG GATCCCTGTG CCCACCAGCA

401 GCACTCCTGG GACCTCCACA GTGGACCTTG GGTCAGGGAC TCCATCCTCC

451 CTCCCCAGCC CCACA

(SEQ ID NO: 87)

1 GCTGCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CCATCACCAA

51 CCTGCAGTAC GAGGAGGACA TGCATCACCC AGGCTCCAGG AAGTTCAACA

101 CCACGGAGCG GGTCCTGCAG GGTCTGCTTG GTCCCATGTT CAAGAACACC

151 AGTGTCGGCC TTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGTCCGA

201 GAAGGATGGA GCAGCCACTG GAGTGGATGC CATCTGCACC CACCGTCTTG

251 ACCCCAAAAG CCCTGGAGTG GACAGGGAGC AGCTATACTG GGAGCTGAGC

301 CAGCTGACCA ATGGCATCAA AGAGCTGGGT CCCTACACCC TGGACAGAAA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCA GACCTCTGCG CCCAACACCA

401 GCACTCCTGG GACCTCCACA GTGGACCTTG GGACCTCAGG GACTCCATCC

451 TCCCTCCCCA GCCCTACA

(SEQ ID NO: 88)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGNG NGTNCTGCAG GGTCTGCTNN NNCCCNTNTT CAAGAACACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGTCCGA

201 GAAGGATGGA GCAGCCACTG GAGTGGATGC CATCTGCACC CACCGTCTTG

251 ACCCCAAAAG CCCTGGAGTG GACAGGGAGC AGCTATACTG GGAGCTGAGC

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

301 CAGCTGACCA ATGGCATCAA AGAGCTGGGT CCCTACACCC TGGACAGAAA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCA GACCTCTGCG CCCAACACCA

401 GCACTCCTGG GACCTCCACA GTGGACCTTG GGACCTCAGG GACTCCATCC

451 TCCCTCCCCA GCCCTACA

(SEQ ID NO: 89)

1 TCTGCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CCATCACCAA

51 CCTGCAGTAC GAGGAGGACA TGCATCACCC AGGCTCCAGG AAGTTCAACA

101 CCACGGAGCG GGTCCTGCAG GGTCTGCTTG GTCCCATGTT CAAGAACACC

151 AGTGTCGGCC TTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGCCTGA

201 GAAGAATGGG GCAGCCACTG GAATGGATGC CATCTGCAGC CACCGTCTTG

251 ACCCCAAAAG CCCTGGACTC AACAGAGAGC AGCTGTACTG GGAGCTGAGC

301 CAGCTGACCC ATGGCATCAA AGAGCTGGGC CCCTACACCC TGGACAGGAA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCG GAGCTCTGTG GCCCCCACCA

401 GCACTCCTGG GACCTCCACA GTGGACCTTG GGACCTCAGG GACTCCATCC

451 TCCCTCCCCA GCCCCACA

(SEQ ID NO: 90)

1 ACAGCTGTTC CTCTCCTGGT GCCGTTCACC CTCAACTTTA CCATCACCAA

51 TCTGCAGTAT GGGGAGGACA TGCGTCACCC TGGCTCCAGG AAGTTCAACA

101 CCACAGAGAG GGTCCTGCAG GGTCTGCTTG GTCCCTTGTT CAAGAACTCC

151 AGTGTCGGCC CTCTGTACTC TGGCTGCAGA CTGATCTCTC TCAGGTCTGA

201 GAAGGATGGG GCAGCCACTG GAGTGGATGC CATCTGCACC CACCACCTTA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

```
      251  ACCCTCAAAG CCCTGGACTG GACAGGGAGC AGCTGTACTG GCAGCTGAGC
      301  CAGATGACCA ATGGCATCAA AGAGCTGGGC CCCTACACCC TGGACCGGAA
      351  CAGTCTCTAC GTCAATGGTT TCACCCATCG GAGCTCTGGG CTCACCACCA
      401  GCACTCCTTG GACTTCCACA GTTGACCTTG GAACCTCAGG GACTCCATCC
      451  CCCGTCCCCA GCCCCACA
```

(SEQ ID NO: 91)

```
        1  ACTGCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CCATCACCAA
       51  CCTGCAGTAT GAGGAGGACA TGCATCGCCC TGGATCTAGG AAGTTCAACA
      101  CCACAGAGAG GGTCCTGCAG GGTCTGCTTA GTCCCATTTT CAAGAACTCC
      151  AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTCTC TCAGGCCCGA
      201  GAAGGATGGG GCAGCAACTG GAATGGATGC TGTCTGCCTC TACCACCCTA
      251  ATCCCAAAAG ACCTGGACTG GACAGAGAGC AGCTGTACTG GGAGCTAAGC
      301  CAGCTGACCC ACAACATCAC TGAGCTGGGC CCCTACAGCC TGGACAGGGA
      351  CAGTCTCTAT GTCAATGGTT TCACCCATCA GAACTCTGTG CCCACCACCA
      401  GTACTCCTGG GACCTCCACA GTGTACTGGG CAACCACTGG GACTCCATCC
      451  TCCTTCCCCG GCCACACA
```

(SEQ ID NO: 92)

```
        1  GAGCCTGGCC CTCTCCTGAT ACCATTCACT TTCAACTTTA CCATCACCAA
       51  CCTGCATTAT GAGGAAAACA TGCAACACCC TGGTTCCAGG AAGTTCAACA
      101  CCACGGAGAG GGTTCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAACACC
      151  AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTCTC TCAGGCCCGA
```

TABLE 15-continued

CA125 Repeat Nucleotide Sequence (SEQ ID NO: 83 thru SEQ ID NO: 145)

201 GAAGGATGGG GCAGCAACTG GAATGGATGC TGTCTGCCTC TACCACCCTA

251 ATCCCAAAAG ACCTGGGCTG GACAGAGAGC AGCTGTACTG GGAGCTAAGC

301 CAGCTGACCC ACAACATCAC TGAGCTGGGC CCCTACAGCC TGGACAGGGA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCA GAACTCTGTG CCCACCACCA

401 GTACTCCTGG GACCTCCACA GTGTACTGGG CAACCACTGG GACTCCATCC

451 TCCTTCCCCG GCCACACA

(SEQ ID NO: 93)

1 GAGCCTGGCC CTCTCCTGAT ACCATTCACT TTCAACTTTA CCATCACCAA

51 CCTGCATTAT GAGGAAAACA TGCAACACCC TGGTTCCAGG AAGTTCAACA

101 CCACGGAGAG GGTTCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAACACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGACCTGA

201 GAAGCATGAG GCAGCCACTG GAGTGGACAC CATCTGTACC CACCGCGTTG

251 ATCCCATCGG ACCTGGACTG GACAGGGAGC GGCTATACTG GGAGCTGAGC

301 CAGCTGACCA ACAGCATTAC CGAACTGGGA CCCTACACCC TGGACAGGGA

351 CAGTCTCTAT GTCAATGGCT TCAACCCTCG GAGCTCTGTG CCAACCACCA

401 GCACTCCTGG GACCTCCACA GTGCACCTGG CAACCTCTGG GACTCCATCC

451 TCCCTGCCTG GCCACACA

(SEQ ID NO: 94)

1 GCCCCTGTCC CTCTCTTGAT ACCATTCACC CTCAACTTTA CCATCACCAA

51 CCTGCATTAT GAGGAAAACA TGCAACACCC TGGTTCCAGG AAGTTCAACA

101 CCACGGAGAG GGTTCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAACACC

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGACCTGA

201 GAAGCATGAG GCAGCCACTG GAGTGGACAC CATCTGTACC CACCGCGTTG

251 ATCCCATCGG ACCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

451 TCCNTCCCCN GCCNCACA

(SEQ ID NO: 95)

1 TCTGCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CCATCACCAA

51 CCTGCAGTAC GAGGAGGACA TGCATCACCC AGGCTCCAGG AAGTTCAACA

101 CCACGGAGCG GGTCCTGCAG GGTCTGCTTG GTCCCATGTT CAAGAACACC

151 AGTGTCGGCC TTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGCCTGA

201 GAAGAATGGG GCAGCCACTG GAATGGATGC CATCTGCAGC CACCGTCTTG

251 ACCCCAAAAG CCCTGGACTC GACAGAGAGC AGCTGTACTG GGAGCTGAGC

301 CAGCTGACCC ATGGCATCAA AGAGCTGGGC CCCTACACCC TGGACAGGAA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCG GAGCTCTGTG GCCCCCACCA

401 GCACTCCTGG GACCTCCACA GTGGACCTTG GGACCTCAGG GACTCCATCC

451 TCCCTCCCCA GCCCCACA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

(SEQ ID NO: 96)

```
  1  ACAGCTGTTC CTCTCCTGGT GCCGTTCACC CTCAACTTTA CCATCACCAA
 51  TCTGCAGTAT GGGGAGGACA TGCGTCACCC TGGCTCCAGG AAGTTCAACA
101  CCACAGAGAG GGTCCTGCAG GGTCTGCTTG GTCCCTTGTT CAAGAACTCC
151  AGTGTCGGCC CTCTGTACTC TGGCTGCAGA CTGATCTCTC TCAGGTCTGA
201  GAAGGATGGG GCAGCCACTG GAGTGGATGC CATCTGCACC CACCACCTTA
251  ACCCTCAAAG CCCTGGACTG GACAGGGAGC AGCTGTACTG GCAGCTGAGC
301  CAGATGACCA ATGGCATCAA AGAGCTGGGC CCCTACACCC TGGACCGGAA
351  CAGTCTCTAC GTCAATGGTT TCACCCATCG GAGCTCTGGG CTCACCACCA
401  GCACTCCTTG GACTTCCACA GTTGACCTTG GAACCTCAGG GACTCCATCC
451  CCCGTCCCCA GCCCCACA
```

(SEQ ID NO: 97)

```
  1  ACTGCTGGCC CTCTCCTGGT GCCATTCACC CTAAACTTCA CCATCACCAA
 51  CCTGCAGTAT GAGGAGGACA TGCATCGCCC TGGATCTAGG AAGTTCAACG
101  CCACAGAGAG GGTCCTGCAG GGTCTGCTTA GTCCCATATT CAAGAACTCC
151  AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTCTC TCAGGCCCGA
201  GAAGGATGGG GCAGCAACTG GAATGGATGC TGTCTGCCTC TACCACCCTA
251  ATCCCAAAAG ACCTGGACTG GACAGAGAGC AGCTGTACTG GGAGCTAAGC
301  CAGCTGACCC ACAACATCAC TGAGCTGGGC CCCTACAGCC TGGACAGGGA
351  CAGTCTCTAT GTCAATGGTT TCACCCATCA GAGCTCTATG ACGACCACCA
```

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

401 GAACTCCTGA TACCTCCACA ATGCACCTGG CAACCTCGAG AACTCCAGCC

451 TCCCTGTCTG GACCTACG

(SEQ ID NO: 98)

1 ACCGCCAGCC CTCTCCTGGT GCTATTCACA ATCAACTGCA CCATCACCAA

51 CCTGCAGTAC GAGGAGGACA TGCGTCGCAC TGGCTCCAGG AAGTTCAACA

101 CCATGGAGAG TGTCCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAACACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA TTGACCTTGC TCAGGCCCAA

201 GAAAGATGGG GCAGCCACTG GAGTGGATGC CATCTGCACC CACCGCCTTG

251 ACCCCAAAAG CCCTGGACTC AACAGGGAGC AGCTGTACTG GGAGCTAAGC

301 AAACTGACCA ATGACATTGA AGAGCTGGGC CCCTACACCC TGGACAGGAA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCA GAGCTCTGTG TCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGGATCTCA GAACCTCAGG GACTCCATCC

451 TCCCTCTCCA GCCCCACAAT TATG

(SEQ ID NO: 99)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGAG GGTCCTACAG GGTCTGCTCA GGCCCTTGTT CAAGAACACC

151 AGTGTCAGCT CTCTGTACTC TGGTTGCAGA CTGACCTTGC TCAGGCCTGA

201 GAAGGATGGG GCAGCCACCA GAGTGGATGC TGCCTGCACC TACCGCCCTG

251 ATCCCAAAAG CCCTGGACTG GACAGAGAGC AACTATACTG GGAGCTGAGC

301 CAGCTAACCC ACAGCATCAC TGAGCTGGGA CCCTACACCC TGGACAGGGT

TABLE 15-continued

CA125 Repeat Nucleotide Sequence (SEQ ID NO: 83 thru 145)

351 CAGTCTCTAT GTCAATGGCT TCAACCCTCG GAGCTCTGTG CCAACCACCA

401 GCACTCCTGG GACCTCCACA GTGCACCTGG CAACCTCTGG GACTCCATCC

451 TCCCTGCCTG GCCACACA

(SEQ ID NO: 100)

1 GCCCCTGTCC CTCTCTTGAT ACCATTCACC CTCAACTTTA CCATCACCAA

51 CCTGCATTAT GAAGAAAACA TGCAACACCC TGGTTCCAGG AAGTTCAACA

101 CCACGGAGAG GGTTCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAGCACC

151 AGCGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGACCTGA

201 GAAACATGGG GCAGCCACTG GAGTGGACGC CATCTGCACC CTCCGCCTTG

251 ATCCCACTGG TCCTGGACTG GACAGAGAGC GGCTATACTG GGAGCTGAGC

301 CAGCTGACCA ACAGCGTTAC AGAGCTGGGC CCCTACACCC TGGACAGGGA

351 CAGTCTCTAT GTCAATGGCT TCACCCAGCG GAGCTCTGTG CCAACCACCA

401 GTATTCCTGG GACCTCTGCA GTGCACCTGG AAACCTCTGG GACTCCAGCC

451 TCCCTCCCTG GCCACACA

(SEQ ID NO: 101)

1 GCCCCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CTATCACCAA

51 CCTGCAGTAT GAGGTGGACA TGCGTCACCC TGGTTCCAGG AAGTTCAACA

101 CCACGGAGAG AGTCCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAGCACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGCCTGA

201 AAAACGTGGG GCAGCCACCG GCGTGGACAC CATCTGCACT CACCGCCTTG

251 ACCCTCTAAA CCCTGGACTG GACAGAGAGC AGCTATACTG GGAGCTGAGC

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

301 AAACTGACCC GTGGCATCAT CGAGCTGGGC CCCTACCTCC TGGACAGAGG

351 CAGTCTCTAT GTCAATGGTT TCACCCATCG GAACTTTGTG CCCATCACCA

401 GCACTCCTGG GACCTCCACA GTACACCTAG GAACCTCTGA AACTCCATCC

451 TCCCTACCTA GACCCATA

(SEQ ID NO: 102)

1 GTGCCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CCATCACCAA

51 CTTGCAGTAT GAGGAGGCCA TGCGACACCC TGGCTCCAGG AAGTTCAATA

101 CCACGGAGAG GGTCCTACAG GGTCTGCTCA GGCCCTTGTT CAAGAATACC

151 AGTATCGGCC CTCTGTACTC CAGCTGCAGA CTGACCTTGC TCAGGCCAGA

201 GAAGGACAAG GCAGCCACCA GAGTGGATGC CATCTGTACC CACCACCCTG

251 ACCCTCAAAG CCCTGGACTG AACAGAGAGC AGCTGTACTG GGAGCTGAGC

301 CAGCTGACCC ACGGCATCAC TGAGCTGGGC CCCTACACCC TGGACAGGGA

351 CAGTCTCTAT GTCGATGGTT TCACTCATTG GAGCCCCATA CCGACCACCA

401 GCACTCCTGG GACCTCCATA GTGAACCTGG GAACCTCTGG GATCCCACCT

451 TCCCTCCCTG AAACTACA

(SEQ ID NO: 103)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGAG GGTTCTGCAG GGTCTGCTCA AACCCTTGTT CAGGAATAGC

151 AGTCTGGAAT ACCTCTATTC AGGCTGCAGA CTAGCCTCAC TCAGGCCAGA

201 GAAGGATAGC TCAGCCATGG CAGTGGATGC CATCTGCACA CATCGCCCTG

TABLE 15-continued

CA125 Repeat Nucleotide Sequence (SEQ ID NO: 83 thru 145)

251 ACCCTGAAGA CCTCGGACTG GACAGAGAGC GACTGTACTG GGAGCTGAGC

301 AATCTGACAA ATGGCATCCA GGAGCTGGGC CCCTACACCC TGGACCGGAA

351 CAGTCTCTAC GTCAATGGTT TCACCCATCG GAGCTCTGGG CTCACCACCA

401 GCACTCCTTG GACTTCCACA GTTGACCTTG GAACCTCAGG GACTCCATCC

451 CCCGTCCCCA GCCCCACA

(SEQ ID NO: 104)

1 ACTGCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CCATCACCAA

51 CCTGCAGTAT GAGGAGGACA TGCATCGCCC TGGTTCCAGG AGGTTCAACA

101 CCACGGAGAG GGTTCTGCAG GGTCTGCTCA CGCCCTTGTT CAAGAACACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGACCTGA

201 GAAGCAAGAG GCAGCCACTG GAGTGGACAC CATCTGTACC CACCGCGTTG

251 ATCCCATCGG ACCTGGACTG GACAGAGAGC GGCTATACTG GGAGCTGAGC

301 CAGCTGACCA ACAGCATCAC AGAGCTGGGA CCCTACACCC TGGATAGGGA

351 CAGTCTCTAT GTCAATGGCT TCAACCCTTG GAGCTCTGTG CCAACCACCA

401 GCACTCCTGG GACCTCCACA GTGCACCTGG CAACCTCTGG GACTCCATCC

451 TCCCTGCCTG GCCACACA

(SEQ ID NO: 105)

1 GCCCCTGTCC CTCTCTTGAT ACCATTCACC CTCAACTTTA CCATCACCGA

51 CCTGCATTAT GAAGAAAACA TGCAACACCC TGGTTCCAGG AAGTTCAACA

101 CCACGGAGAG GGTTCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAGCACC

151 AGCGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGACCTGA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

201 GAAACATGGG GCAGCCACTG GAGTGGACGC CATCTGCACC CTCCGCCTTG

251 ATCCCACTGG TCCTGGACTG GACAGAGAGC GGCTATACTG GGAGCTGAGC

301 CAGCTGACCA ACAGCGTTAC AGAGCTGGGC CCCTACACCC TGGACAGGGA

351 CAGTCTCTAT GTCAATGGCT TCACCCATCG GAGCTCTGTG CCAACCACCA

401 GTATTCCTGG GACCTCTGCA GTGCACCTGG AAACCTCTGG GACTCCAGCC

451 TCCCTCCCTG GCCACACA

(SEQ ID NO: 106)

1 GCCCCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CTATCACCAA

51 CCTGCAGTAT GAGGAGGACA TGCGTCACCC TGGTTCCAGG AAGTTCAGCA

101 CCACGGAGAG AGTCCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAACACC

151 AGTGTCAGCT CTCTGTACTC TGGTTGCAGA CTGACCTTGC TCAGGCCTGA

201 GAAGGATGGG GCAGCCACCA GAGTGGATGC TGTCTGCACC CATCGTCCTG

251 ACCCCAAAAG CCCTGGACTG GACAGAGAGC GGCTGTACTG GAAGCTGAGC

301 CAGCTGACCC ACGGCATCAC TGAGCTGGGC CCCTACACCC TGGACAGGCA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCA GAGCTCTATG ACGACCACCA

401 GAACTCCTGA TACCTCCACA ATGCACCTGG CAACCTCGAG AACTCCAGCC

451 TCCCTGTCTG GACCTACG

(SEQ ID NO: 107)

1 ACCGCCAGCC CTCTCCTGGT GCTATTCACA ATTAACTTCA CCATCACTAA

51 CCTGCGGTAT GAGGAGAACA TGCATCACCC TGGCTCTAGA AAGTTTAACA

101 CCACGGAGAG AGTCCTTCAG GGTCTGCTCA GGCCTGTGTT CAAGAACACC

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCACGC TCAGGCCCAA

201 GAAGGATGGG GCAGCCACCA AAGTGGATGC CATCTGCACC TACCGCCCTG

251 ATCCCAAAAG CCCTGGACTG GACAGAGAGC AGCTATACTG GGAGCTGAGC

301 CAGCTAACCC ACAGCATCAC TGAGCTGGGC CCCTACACCC AGGACAGGGA

351 CAGTCTCTAT GTCAATGGCT TCACCCATCG GAGCTCTGTG CCAACCACCA

401 GTATTCCTGG GACCTCTGCA GTGCACCTGG AAACCTCTGG GACTCCAGCC

451 TCCCTCCCTG GCCACACA

(SEQ ID NO: 108)

1 GCCCCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CTATCACCAA

51 CCTGCAGTAT GAGGAGGACA TGCGTCACCC TGGTTCCAGG AAGTTCAACA

101 CCACGGAGAG AGTCCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAGCACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGCCTGA

201 AAAACGTGGG GCAGCCACCG GCGTGGACAC CATCTGCACT CACCGCCTTG

251 ACCCTCTAAA CCCAGGACTG GACAGAGAGC AGCTATACTG GGAGCTGAGC

301 AAACTGACCC GTGGCATCAT CGAGCTGGGC CCCTACCTCC TGGACAGAGG

351 CAGTCTCTAT GTCAATGGTT TCACCCATCG GACCTCTGTG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGGACCTTG GAACCTCAGG GACTCCATTC

451 TCCCTCCCAA GCCCCGCA

(SEQ ID NO: 109)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence (SEQ ID NO: 83 thru SEQ ID NO: 145)

101 CCACNGAGAG GGTCCTGCAG ACTCTGCTTG GTCCTATGTT CAAGAACACC

151 AGTGTTGGCC TTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGTCCGA

201 GAAGGATGGA GCAGCCACTG GAGTGGATGC CATCTGCACC CACCGTCTTG

251 ACCCCAAAAG CCCTGGAGTG GACAGGGAGC AACTATACTG GGAGCTGAGC

301 CAGCTGACCA ATGGCATTAA AGAACTGGGC CCCTACACCC TGGACAGGAA

351 CAGTCTCTAT GTCAATGGGT TCACCCATTG GATCCCTGTG CCCACCAGCA

401 GCACTCCTGG GACCTCCACA GTGGACCTTG GGTCAGGGAC TCCATCCTCC

451 CTCCCCAGCC CCACA

(SEQ ID NO: 110)

1 ACTGCTGGCC CTCTCCTGGT GCCGTTCACC CTCAACTTCA CCATCACCAA

51 CCTGAAGTAC GAGGAGGACA TGCATTGCCC TGGCTCCAGG AAGTTCAACA

101 CCACAGAGAG AGTCCTGCAG AGTCTGCTTG GTCCCATGTT CAAGAACACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGTCCGA

201 GAAGGATGGA GCAGCCACTG GAGTGGATGC CATCTGCACC CACCGTCTTG

251 ACCCCAAAAG CCCTGGAGTG GACAGGGAGC AGCTATACTG GGAGCTGAGC

301 CAGCTGACCA ATGGCATCAA AGAGCTGGGT CCCTACACCC TGGACAGAAA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCA GACCTCTGCG CCCAACACCA

401 GCACTCCTGG GACCTCCACA GTGGACCTTG GGACCTCAGG GACTCCATCC

451 TCCCTCCCCA GCCCTACA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

(SEQ ID NO: 111)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGNG NGTNCTGCAG GGTCTGCTNN NNCCCNTNTT CAAGAACNCC

151 AGTGTNGGCC NTCTGTACTC TGGCTGCAGA CTGACCTNNC TCAGGNCNGA

201 GAAGNATGGN GCAGCCACTG GANTGGATGC CATCTGCANC CACCNNCNTN

251 ANCCCAAAAG NCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATTG GATCCCTGTG CCCACCAGCA

401 GCACTCCTGG GACCTCCACA GTGGACCTTG GGTCAGGGAC TCCATCCTCC

451 CTCCCCAGCC CCACA

(SEQ ID NO: 112)

1 ACTGCTGGCC CTCTCCTGGT GCCGTTCACC CTCAACTTCA CCATCACCAA

51 CCTGAAGTAC GAGGAGGACA TGCATTGCCC TGGCTCCAGG AAGTTCAACA

101 CCACAGAGAG AGTCCTGCAG AGTCTGCTTG GTCCCATGTT CAAGAACACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTCGC TCAGGTCCGA

201 GAAGGATGGA GCAGCCACTG GAGTGGATGC CATCTGCACC CACCGTGTTG

251 ACCCCAAAAG CCCTGGAGTG GACAGGGAGC AGCTATACTG GGAGCTGAGC

301 CAGCTGACCA ATGGCATCAA AGAGCTGGGT CCCTACACCC TGGACAGAAA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCA GACCTCTGCG CCCAACACCA

401 GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

451 TCCNTCCCCN GCCNCACA

(SEQ ID NO: 113)

1 TCTGCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CCATCACCAA

51 CCTGCAGTAC GAGGAGGACA TGCATCACCC AGGCTCCAGG AAGTTCAACA

101 CCACGGAGCG GGTCCTGCAG GGTCTGCTTG GTCCCATGTT CAAGAACACC

151 AGTGTCGGCC TTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGCCTGA

201 GAAGAATGGG GCAACCACTG GAATGGATGC CATCTGCACC CACCGTCTTG

251 ACCCCAAAAG CCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

451 TCCNTCCCCN GCCNCACA

(SEQ ID NO: 114)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGAG GGTTCTGCAG GGTCTGCTCA AACCCTTGTT CAGGAATAGC

151 AGTCTGGAAT ACCTCTATTC AGGCTGCAGA CTAGCCTCAC TCAGGCCAGA

201 GAAGGATAGC TCAGCCATGG CAGTGGATGC CATCTGCACA CATCGCCCTG

251 ACCCTGAAGA CCTCGGACTG GACAGAGAGC GACTGTACTG GGAGCTGAGC

301 AATCTGACAA ATGGCATCCA GGAGCTGGGC CCCTACACCC TGGACCGGAA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCG AAGCTCTATG CCCACCACCA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

401 GCACTCCTGG GACCTCCACA GTGGATGTGG GAACCTCAGG GACTCCATCC

451 TCCAGCCCCA GCCCCACG

(SEQ ID NO: 115)

1 ACTGCTGGCC CTCTCCTGAT ACCATTCACC CTCAACTTCA CCATCACCAA

51 CCTGCAGTAT GGGGAGGACA TGGGTCACCC TGGCTCCAGG AAGTTCAACA

101 CCACAGAGAG GGTCCTGCAG GGTCTGCTTG GTCCCATATT CAAGAACACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTCTC TCAGGTCTGA

201 GAAGGATGGA GCAGCCACTG GAGTGGATGC CATCTGCATC CATCATCTTG

251 ACCCCAAAAG CCCTGGACTC AACAGAGAGC GGCTGTACTG GGAGCTGAGC

301 CAACTGACCA ATGGCATCAA AGAGCTGGGC CCCTACACCC TGGACAGGAA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCG GACCTCTGTG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGGACCTTG GAACCTCAGG GACTCCATTC

451 TCCCTCCCAA GCCCCGCA

(SEQ ID NO: 116)

1 ACTGCTGGCC CTCTCCTGGT GCTGTTCACC CTCAACTTCA CCATCACCAA

51 CCTGAAGTAT GAGGAGGACA TGCATCGCCC TGGCTCCAGG AAGTTCAACA

101 CCACTGAGAG GGTCCTGCAG ACTCTGCTTG GTCCTATGTT CAAGAACACC

151 AGTGTTGGCC TTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGTCCGA

201 GAAGGATGGA GCAGCCACTG GAGTGGATGC CATCTGCACC CACCGTCTTG

251 ACCCCAAAAG CCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

451 TCCNTCCCCN GCCNCACA

(SEQ ID NO: 117)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGAG AGTCCTTCAG GGTCTGCTCA GGCCTGTGTT CAAGAACACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGCCCAA

201 GAAGGATGGG GCAGCCACCA AAGTGGATGC CATCTGCACC TACCGCCCTG

251 ATCCCAAAAG CCCTGGACTG GACAGAGAGC AGCTATACTG GGAGCTGAGC

301 CAGCTAACCC ACAGCATCAC TGAGCTGGGC CCCTACACCC AGGACAGGGA

351 CAGTCTCTAT GTCAATGGCT TCACCCATCG GAGCTCTGTG CCAACCACCA

401 GTATTCCTGG GACCTCTGCA GTGCACCTGG AAACCACTGG GACTCCATCC

451 TCCTTCCCCG GCCACACA

(SEQ ID NO: 118)

1 GAGCCTGGCC CTCTCCTGAT ACCATTCACT TTCAACTTTA CCATCACCAA

51 CCTGCGTTAT GAGGAAAACA TGCAACACCC TGGTTCCAGG AAGTTCAACA

101 CCACGGAGAG GGTTCTGCAG GGTCTGCTCA CGCCCTTGTT CAAGAACACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGACCTGA

201 GAAGCAGGAG GCAGCCACTG GAGTGGACAC CATCTGTACC CACCGCGTTG

TABLE 15-continued

CA125 Repeat Nucleotide Sequence (SEQ ID NO: 83 thru 145)

251 ATCCCATCGG ACCTGGACTG GACAGAGAGC GGCTATACTG GGAGCTGAGC

301 CAGCTGACCA ACAGCATCAC AGAGCTGGGA CCCTACACCC TGGATAGGGA

351 CAGTCTCTAT GTCGATGGCT TCAACCCTTG GAGCTCTGTG CCAACCACCA

401 GCACTCCTGG GACCTCCACA GTGCACCTGG CAACCTCTGG GACTCCATCC

451 CCCCTGCCTG GCCACACA

(SEQ ID NO: 119)

1 GCCCCTGTCC CTCTCTTGAT ACCATTCACC CTCAACTTTA CCATCACCGA

51 CCTGCATTAT GAAGAAAACA TGCAACACCC TGGTTCCAGG AAGTTCAACA

101 CCACGGAGAG GGTTCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAGCACC

151 AGCGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGACCTGA

201 GAAACATGGG GCAGCCACTG GAGTGGACGC CATCTGCACC CTCCGCCTTG

251 ATCCCACTGG TCCTGGACTG GACAGAGAGC GGCTATACTG GGAGCTGAGC

301 CAGCTGACCA ACAGCATCAC AGAGCTGGGA CCCTACACCC TGGATAGGGA

351 CAGTCTCTAT GTCAATGGCT TCAACCCTTG GAGCTCTGTG CCAACCACCA

401 GCACTCCTGG GACCTCCACA GTGCACCTGG CAACCTCTGG GACTCCATCC

451 TCCCTGCCTG GCCACACA

(SEQ ID NO: 120)

1 ACTGCTGGCC CTCTCCTGGT GCCGTTCACC CTCAACTTCA CCATCACCAA

51 CCTGAAGTAC GAGGAGGACA TGCATTGCCC TGGCTCCAGG AAGTTCAACA

101 CCACAGAGAG AGTCCTGCAG AGTCTGCATG GTCCCATGTT CAAGAACACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGTCCGA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

201 GAAGGATGGA GCAGCCACTG GAGTGGATGC CATCTGCACC CACCGTCTTG

251 ACCCCAAAAG CCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

451 TCCNTCCCCN GCCNCACA

(SEQ ID NO: 121)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGNG NGTNCTGCAG GGTCTGCTNN NNCCCNTNTT CAAGAACNCC

151 AGTGTNGGCC NTCTGTACTC TGGCTGCAGA CTGACCTNNC TCAGGNCNGA

201 GAAGNATGGN GCAGCCACTG GANTGGATGC CATCTGCANC CACCNNCNTN

251 ANCCCAAAAG NCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ACAGCATCAC AGAGCTGGGA CCCTACACCC TGGATAGGGA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCG AAGCTCTATG CCCACCACCA

401 GTATTCCTGG GACCTCTGCA GTGCACCTGG AAACCTCTGG GACTCCAGCC

451 TCCCTCCCTG GCCACACA

(SEQ ID NO: 122)

1 GCCCCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CTATCACCAA

51 CCTGCAGTAT GAGGAGGACA TGCGTCACCC TGGTTCCAGG AAGTTCAACA

101 CCACGGAGAG AGTCCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAGCACC

TABLE 15-continued

CA125 Repeat Nucleotide Sequence (SEQ ID NO: 83 thru SEQ ID NO: 145)

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGCCTGA

201 AAAACGTGGG GCAGCCACCG GCGTGGACAC CATCTGCACT CACCGCCTTG

251 ACCCTCTAAA CCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

451 TCCNTCCCCN GCCNCACA

(SEQ ID NO: 123)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGNG NGTNCTGCAG GGTCTGCTNN NNCCCNTNTT CAAGAACNCC

151 AGTGTNGGCC NTCTGTACTC TGGCTGCAGA CTGACCTNNC TCAGGNCNGA

201 GAAGNATGGN GCAGCCACTG GANTGGATGC CATCTGCANC CACCNNCNTN

251 ANCCCAAAAG NCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TTCACCCTCG GAGCTCTGTG CCAACCACCA

401 GCACTCCTGG GACCTCCACA GTGCACCTGG CAACCTCTGG GACTCCATCC

451 TCCCTGCCTG GCCACACA

(SEQ ID NO: 124)

1 GCCCCTGTCC CTCTCTTGAT ACCATTCACC CTCAACTTTA CCATCACCAA

51 CCTGCATTAT GAAGAAAACA TGCAACACCC TGGTTCCAGG AAGTTCAACA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

101 CCACGGAGCG GGTCCTGCAG GGTCTGCTTG GTCCCATGTT CAAGAACACA

151 AGTGTCGGCC TTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGCCTGA

201 GAAGAATGGG GCAGCCACTG GAATGGATGC CATCTGCAGC CACCGTCTTG

251 ACCCCAAAAG CCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

451 TCCNTCCCCN GCCNCACA

(SEQ ID NO: 125)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGNG NGTNCTGCAG GGTCTGCTNN NNCCCNTNTT CAAGAACNCC

151 AGTGTNGGCC NTCTGTACTC TGGCTGCAGA CTGACCTNNC TCAGGNCNGA

201 GAAGNATGGN GCAGCCACTG GANTGGATGC CATCTGCANC CACCNNCNTN

251 ANCCCAAAAG NCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCA GAACTCTGTG CCCACCACCA

401 GTACTCCTGG GACCTCCACA GTGTACTGGG CAACCACTGG GACTCCATCC

451 TCCTTCCCCG GCCACACA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

(SEQ ID NO: 126)

1 GAGCCTGGCC CTCTCCTGAT ACCATTCACT TTCAACTTTA CCATCACCAA

51 CCTGCATTAT GAGGAAAACA TGCAACACCC TGGTTCCAGG AAGTTCAACA

101 CCACGGAGAG GGTTCTGCAG GGTCTGCTCA CGCCCTTGTT CAAGAACACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGACCTGA

201 GAAGCAGGAG GCAGCCACTG GAGTGGACAC CATCTGTACC CACCGCGTTG

251 ATCCCATCGG ACCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

451 TCCNTCCCCN GCCNCACA

(SEQ ID NO: 127)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGNG NGTNCTGCAG GGTCTGCTNN NNCCCNTNTT CAAGAACNCC

151 AGTGTNGGCC NTCTGTACTC TGGCTGCAGA CTGACCTNNC TCAGGNCNGA

201 GAAGNATGGN GCAGCCACTG GANTGGATGC CATCTGCANC CACCNNCNTN

251 ANCCCAAAAG NCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCG GAGCTCTGTG CCAACCACCA

401 GCAGTCCTGG GACCTCCACA GTGCACCTGG CAACCTCTGG GACTCCATCC

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

451 TCCCTGCCTG GCCACACA

(SEQ ID NO: 128)

1 GCCCCTGTCC CTCTCTTGAT ACCATTCACC CTCAACTTTA CCATCACCAA

51 CCTGCATTAT GAAGAAAACA TGCAACACCC TGGTTCCAGG AAGTTCAACA

101 CCACGGAGAG GGTTCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAGCACC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGACCTGA

201 GAAACATGGG GCAGCCACTG GAGTGGACGC CATCTGCACC CTCCGCCTTG

251 ATCCCACTGG TCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

451 TCCNTCCCCN GCCNCACA

(SEQ ID NO: 129)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGNG NGTNCTGCAG GGTCTGCTNN NNCCCNTNTT CAAGAACNCC

151 AGTGTNGGCC NTCTGTACTC TGGCTGCAGA CTGACCTNNC TCAGGNCNGA

201 GAAGNATGGN GCAGCCACTG GANTGGATGC CATCTGCANC CACCNNCNTN

251 ANCCCAAAAG NCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCG GACCTCTGTG CCCACCACCA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence (SEQ ID NO: 83 thru SEQ ID NO: 145)

401 GCACTCCTGG GACCTCCACA GTGCACCTGG CAACCTCTGG GACTCCATCC

451 TCCCTGCCTG GCCACACA

(SEQ ID NO: 130)

1 GCCCCTGTCC CTCTCTTGAT ACCATTCACC CTCAACTTTA CCATCACCAA

51 CCTGCAGTAT GAGGAGGACA TGCATCGCCC TGGATCTAGG AAGTTCAACA

101 CCACAGAGAG GGTCCTGCAG GGTCTGCTTA GTCCCATTTT CAAGAACTCC

151 AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTCTC TCAGGCCCGA

201 GAAGGATGGG GCAGCAACTG GAATGGATGC TGTCTGCCTC TACCACCCTA

251 ATCCCAAAAG ACCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

451 TCCNTCCCCN GCCNCACA

(SEQ ID NO: 131)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGNG NGTNCTGCAG GGTCTGCTNN NNCCCNTNTT CAAGAACNCC

151 AGTGTNGGCC NTCTGTACTC TGGCTGCAGA CTGACCTNNC TCAGGNCNGA

201 GAAGNATGGN GCAGCCACTG GANTGGATGC CATCTGCANC CACCNNCNTN

251 ANCCCAAAAG NCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

```
351  CAGTCTCTAT GTCAATGGTT TCACCCATTG GAGCTCTGGG CTCACCACCA

401  GCACTCCTTG GACTTCCACA GTTGACCTTG GAACCTCAGG GACTCCATCC

451  CCCGTCCCCA GCCCCACA
```

(SEQ ID NO: 132)

```
1    ACTGCTGGCC CTCTCCTGGT GCCATTCACC CTAAACTTCA CCATCACCAA

51   CCTGCAGTAT GAGGAGGACA TGCATCGCCC TGGATCTAGG AAGTTCAACG

101  CCACAGAGAG GGTCCTGCAG GGTCTGCTTA GTCCCATATT CAAGAACACC

151  AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGACCTGA

201  GAAGCAGGAG GCAGCCACTG GAGTGGACAC CATCTGTACC CACCGCGTTG

251  ATCCCATCGG ACCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301  CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351  CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA

401  GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

451  TCCNTCCCCN GCCNCACA
```

(SEQ ID NO: 133)

```
1    NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51   CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101  CCACNGAGNG NGTNCTGCAG GGTCTGCTNN NNCCCNTNTT CAAGAACNCC

151  AGTGTNGGCC NTCTGTACTC TGGCTGCAGA CTGACCTNNC TCAGGNCNGA

201  GAAGNATGGN GCAGCCACTG GANTGGATGC CATCTGCANC CACCNNCNTN

251  ANCCCAAAAG NCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC
```

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCG GAGCTTTGGG CTCACCACCA

401 GCACTCCTTG GACTTCCACA GTTGACCTTG GAACCTCAGG GACTCCATCC

451 CCCGTCCCCA GCCCCACA

(SEQ ID NO: 134)

1 ACTGCTGGCC CTCTCCTGGT GCCATTCACC CTAAACTTCA CCATCACCAA

51 CCTGCAGTAT GAGGAGGACA TGCATCGCCC TGGCTCCAGG AAGTTCAACA

101 CCACGGAGAG GGTCCTTCAG GGTCTGCTTA CGCCCTTGTT CAGGAACACC

151 AGTGTCAGCT CTCTGTACTC TGGTTGCAGA CTGACCTTGC TCAGGCCTGA

201 GAAGGATGGG GCAGCCACCA GAGTGGATGC TGTCTGCACC CATCGTCCTG

251 ACCCCAAAAG CCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

451 TCCNTCCCCN GCCNCACA

(SEQ ID NO: 135)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGNG NGTNCTGCAG GGTCTGCTNN NNCCCNTNTT CAAGAACNCC

151 AGTGTNGGCC NTCTGTACTC TGGCTGCAGA CTGACCTNNC TCAGGNCNGA

201 GAAGNATGGN GCAGCCACTG GANTGGATGC CATCTGCANC CACCNNCNTN

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

```
  251  ANCCCAAAAG NCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC
  301  CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA
  351  CAGTCTCTAT GTCAATGGTT TCACCCATTG GATCCCTGTG CCCACCAGCA
  401  GCACTCCTGG GACCTCCACA GTGGACCTTG GGTCAGGGAC TCCATCCTCC
  451  CTCCCCAGCC CCACA
```

(SEQ ID NO: 136)

```
    1  ACTGCTGGCC CTCTCCTGGT ACCATTCACC CTCAACTTCA CCATCACCAA
   51  CCTGCAGTAT GGGGAGGACA TGGGTCACCC TGGCTCCAGG AAGTTCAACA
  101  CCACAGAGAG GGTCCTGCAG GGTCTGCTTG GTCCCATATT CAAGAACACC
  151  AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTCTC TCAGGTCCGA
  201  GAAGGATGGA GCAGCCACTG GAGTGGATGC CATCTGCATC CATCATCTTG
  251  ACCCCAAAAG CCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC
  301  CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA
  351  CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA
  401  GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC
  451  TCCNTCCCCN GCCNCACA
```

(SEQ ID NO: 137)

```
    1  NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA
   51  CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA
  101  CCACNGAGNG NGTNCTGCAG GGTCTGCTNN NNCCCNTNTT CAAGAACNCC
  151  AGTGTNGGCC NTCTGTACTC TGGCTGCAGA CTGACCTNNC TCAGGNCNGA
```

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

201 GAAGNATGGN GCAGCCACTG GANTGGATGC CATCTGCANC CACCNNCNTN

251 ANCCCAAAAG NCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCA GACCTTTGCG CCCAACACCA

401 GCACTCCTGG GACCTCCACA GTGGACCTTG GGACCTCAGG GACTCCATCC

451 TCCCTCCCC AGCCCTACA

(SEQ ID NO: 138)

1 TCTGCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CCATCACCAA

51 CCTGCAGTAC GAGGAGGACA TGCATCACCC AGGCTCCAGG AAGTTCAACA

101 CCACGGAGCG GGTCCTGCAG GGTCTGCTTG GTCCCATGTT CAAGAACACC

151 AGTGTCGGCC TTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGCCTGA

201 GAAGAATGGG GCAGCCACCA GAGTGGATGC TGTCTGCACC CATCGTCCTG

251 ACCCCAAAAG CCCTGGACTG NACAGNGAGC NGCTNTACTG GGAGCTNAGC

301 CANCTGACCA ANNNCATCNN NGAGCTGGGN CCCTACACCC TGGACAGGNA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCN GANCTCTGNG CCCACCACCA

401 GCACTCCTGG GACCTCCACA GTGNACNTNG GNACCTCNGG GACTCCATCC

451 TCCNTCCCCN GCCNCACA

(SEQ ID NO: 139)

1 NCNNCTGNCC CTCTCCTGNT NCCNTTCACC NTCAACTTNA CCATCACCAA

51 CCTGCANTAN GNGGANNACA TGCNNCNCCC NGGNTCCAGG AAGTTCAACA

101 CCACNGAGAG GGTTCTGCAG GGTCTGCTCA AGCCCTTGTT CAAGAGCACC

TABLE 15-continued

CA125 Repeat Nucleotide Sequence (SEQ ID NO: 83 thru SEQ ID NO: 145)

151 AGTGTTGGCC CTCTGTATTC TGGCTGCAGA CTGACCTTGC TCAGGCCTGA

201 GAAGGACGGA GTAGCCACCA GAGTGGACGC CATCTGCACC CACCGCCCTG

251 ACCCCAAAAT CCCTGGGCTA GACAGACAGC AGCTATACTG GGAGCTGAGC

301 CAGCTGACCC ACAGCATCAC TGAGCTGGGA CCCTACACCC TGGATAGGGA

351 CAGTCTCTAT GTCAATGGTT TCACCCAGCG GAGCTCTGTG CCCACCACCA

401 GCACTCCTGG GACTTTCACA GTACAGCCGG AAACCTCTGA GACTCCATCA

451 TCCCTCCCTG GCCCCACA

(SEQ ID NO: 140)

1 GCCACTGGCC CTGTCCTGCT GCCATTCACC CTCAATTTTA CCATCACTAA

51 CCTGCAGTAT GAGGAGGACA TGCATCGCCC TGGCTCCAGG AAGTTCAACA

101 CCACGGAGAG GGTCCTTCAG GGTCTGCTTA TGCCCTTGTT CAAGAACACC

151 AGTGTCAGCT CTCTGTACTC TGGTTGCAGA CTGACCTTGC TCAGGCCTGA

201 GAAGGATGGG GCAGCCACCA GAGTGGATGC TGTCTGCACC CATCGTCCTG

251 ACCCCAAAAG CCCTGGACTG GACAGAGAGC GGCTGTACTG GAAGCTGAGC

301 CAGCTGACCC ACGGCATCAC TGAGCTGGGC CCCTACACCC TGGACAGGCA

351 CAGTCTCTAT GTCAATGGTT TCACCCATCA GAGCTCTATG ACGACCACCA

401 GAACTCCTGA TACCTCCACA ATGCACCTGG CAACCTCGAG AACTCCAGCC

451 TCCCTGTCTG GACCTACG

(SEQ ID NO: 141)

1 ACCGCCAGCC CTCTCCTGGT GCTATTCACA ATTAACTTCA CCATCACTAA

51 CCTGCGGTAT GAGGAGAACA TGCATCACCC TGGCTCTAGA AAGTTTAACA

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

```
  101  CCACGGAGAG AGTCCTTCAG GGTCTGCTCA GGCCTGTGTT CAAGAACACC
  151  AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTTGC TCAGGCCCAA
  201  GAAGGATGGG GCAGCCACCA AAGTGGATGC CATCTGCACC TACCGCCCTG
  251  ATCCCAAAAG CCCTGGACTG GACAGAGAGC AGCTATACTG GGAGCTGAGC
  301  CAGCTAACCC ACAGCATCAC TGAGCTGGGC CCCTACACCC TGGACAGGGA
  351  CAGTCTCTAT GTCAATGGTT TCACACAGCG GAGCTCTGTG CCCACCACTA
  401  GCATTCCTGG GACCCCCACA GTGGACCTGG GAACATCTGG GACTCCAGTT
  451  TCTAAACCTG GTCCCTCG
```

(SEQ ID NO: 142)

```
    1  GCTGCCAGCC CTCTCCTGGT GCTATTCACT CTCAACTTCA CCATCACCAA
   51  CCTGCGGTAT GAGGAGAACA TGCAGCACCC TGGCTCCAGG AAGTTCAACA
  101  CCACGGAGAG GGTCCTTCAG GGCCTGCTCA GGTCCCTGTT CAAGAGCACC
  151  AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACTTTGC TCAGGCCTGA
  201  AAAGGATGGG ACAGCCACTG GAGTGGATGC CATCTGCACC CACCACCCTG
  251  ACCCCAAAAG CCCTAGGCTG GACAGAGAGC AGCTGTATTG GGAGCTGAGC
  301  CAGCTGACCC ACAATATCAC TGAGCTGGGC CACTATGCCC TGGACAACGA
  351  CAGCCTCTTT GTCAATGGTT TCACTCATCG GAGCTCTGTG TCCACCACCA
  401  GCACTCCTGG GACCCCCACA GTGTATCTGG GAGCATCTAA GACTCCAGCC
  451  TCGATATTTG GCCCTTCA
```

TABLE 15-continued

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 83 thru SEQ ID NO: 145)

(SEQ ID NO: 143)

```
   1  GCTGCCAGCC ATCTCCTGAT ACTATTCACC CTCAACTTCA CCATCACTAA
  51  CCTGCGGTAT GAGGAGAACA TGTGGCCTGG CTCCAGGAAG TTCAACACTA
 101  CAGAGAGGGT CCTTCAGGGC CTGCTAAGGC CCTTGTTCAA GAACACCAGT
 151  GTTGGCCCTC TGTACTCTGG CTCCAGGCTG ACCTTGCTCA GGCCAGAGAA
 201  AGATGGGGAA GCCACCGGAG TGGATGCCAT CTGCACCCAC CGCCCTGACC
 251  CCACAGGCCC TGGGCTGGAC AGAGAGCAGC TGTATTTGGA GCTGAGCCAG
 301  CTGACCCACA GCATCACTGA GCTGGGCCCC TACACACTGG ACAGGGACAG
 351  TCTCTATGTC AATGGTTTCA CCCATCGGAG CTCTGTACCC ACCACCAGC
```

(SEQ ID NO: 144)

```
   1  ACCGGGGTGG TCAGCGAGGA GCCATTCACA CTGAACTTCA CCATCAACAA
  51  CCTGCGCTAC ATGGCGGACA TGGGCCAACC CGGCTCCCTC AAGTTCAACA
 101  TCACAGACAA CGTCATGAAG CACCTGCTCA GTCCTTTGTT CCAGAGGAGC
 151  AGCCTGGGTG CACGGTACAC AGGCTGCAGG GTCATCGCAC TAAGGTCTGT
 201  GAAGAACGGT GCTGAGACAC GGGTGGACCT CCTCTGCACC TACCTGCAGC
 251  CCCTCAGCGG CCCAGGTCTG CCTATCAAGC AGGTGTTCCA TGAGCTGAGC
 301  CAGCAGACCC ATGGCATCAC CCGGCTGGGC CCCTACTCTC TGGACAAAGA
 351  CAGCCTCTAC CTTAACGGTT ACAATGAACC TGGTCTAGAT GAGCCTCCTA
 401  CAACTCCCAA GCCAGCCACC ACATTCCTGC CTCCTCTGTC AGAAGCCACA
 451  ACA
```

TABLE 15-continued

CA125 Repeat Nucleotide Sequence (SEQ ID NO: 83 thru SEQ ID NO: 145)

(SEQ ID NO: 145)

```
   1  GCCATGGGGT ACCACCTGAA GACCCTCACA CTCAACTTCA CCATCTCCAA

  51  TCTCCAGTAT TCACCAGATA TGGGCAAGGG CTCAGCTACA TTCAACTCCA

 101  CCGAGGGGGT CCTTCAGCAC CTGCTCAGAC CCTTGTTCCA GAAGAGCAGC

 151  ATGGGCCCCT TCTACTTGGG TTGCCAACTG ATCTCCCTCA GGCCTGAGAA

 201  GGATGGGGCA GCCACTGGTG TGGACACCAC CTGCACCTAC CACCCTGACC

 251  CTGTGGGCCC CGGGCTGGAC ATACAGCAGC TTTACTGGGA GCTGAGTCAG

 301  CTGACCCATG GTGTCACCCA ACTGGGCTTC TATGTCCTGG ACAGGGATAG

 351  CCTCTTCATC AATGGCTATG CACCCCAGAA TTTATCAATC CGGGGCGAGT

 401  ACCAGATAAA TTTCCACATT GTCAACTGGA ACCTCAGTAA TCCAGACCCC

 451  ACATCCTCAG AGTAC
```

TABLE 16

CA125 Repeat Domains
(SEQ ID NO: 146)

ATVPFMVPFTLNFTITNLQYEEDMRHPGSRKFNATERELQGLLKPLFRNSSLEYLYSGCRLASLRPEKDSSAMAVDAICTHRPDPEDLGLDRERLYWELSNLTNGIQELGPYTLDRNSLYVNGFTHRSSMPTTSTPGTSTVDVGTSGTPSSSPSPT
AAGPLLMPFTLNFTITNLQYEEDMRRTGSRKFNTMESVLQGLLKPLFKNTSVGPLYSGCRLTLLRPEKDGAATGVDAICIHHLDPKSPGLNREQLYWELSKLTNDIEELGPYTLDRNSLYVNGFTHQSSVSTTSTPGTSTVDLRTSGTPSSLSSPTIM
AAGPLLVPFTLNFTITNLQYGEDMGHPGSRKFNTTERVLQGLLGPIFKNTSVGPLYSGCRLTLLRSEKDGAATGVDAICTHRLDPKSPGLDREQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHWIPVPTSSTPGTSTVDLG.SGTPSSLPSPT
TAGPLLVLFTLNFTITNLKYEEDMHRPGSRKFNTTERVLQTLLGPMFKNTSVGLLYSGCRLTLLRSEKDGAATGVDAICTHRLDPKSPGVDREQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHQTSAPNTSTPGTSTVDLGTSGTPSSLPSPT
AAGPLLVPFTLNFTITNLQYEEDMHHPGSRKFNTTERVLQGLLKPLFKSTSVGPLYSGCRLTLLRSEKDGAATGVDAICTHRLDPKSPGVDREQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHRSSGLTTSTPWTSTVDLGTSGTPSSLPSPT
SAGPLLVPFTLNFTITNLQYEEDMRHPGSRKFNTTERVLQGLLGPMFKNTSVGLLYSGCRLTLLRPEKNGAATGMDAICSHRLDPKSPGLNREQLYWELSQLTHGIKELGPYTLDRNSLYVNGFTHRSSGLTTSTPWTSTVDLGTSGTPSPVPSPT
SAGPLLVPFTLNFTITNLQYEEDMHHPGSRKFNTTERVLQGLLGPMFKNTSVGLLYSGCRLTLLRPEKNGAATGMDAVCLYHPNPKRPGLDREQLYWQLSQMTNGIKELGPYTLDRNSLYVNGFTHQNSVPTTSTPGTSTVYWATTGTPSSFPGHT
TAVPLLVPFTLNFTITNLQYGEDMRHPGSRKFNATERVLQGLLSPIFKNSSVGPLYSGCRLTSLRPEKDGAATGMDAVCLYHPNPKRPGLDREQLYCELSQLTHNITELGPYSLDRDSLYVNGFNPRSSVPTTSTPGTSTVVHLATSGTPSSLPGHT
TAGPLLVPFTLNFTITNLQYEEDMHRPGSRKFNTTERVLQGLLKPLFKNTSVGPLYSGCRLTSLRPEKDGAATGMDAVCLYHPNPKRPGLDREQLYCELSQLTHNITELGPYSLDRDSLYVNGFNPRSSVPTTSTPGTSTVVHLATSGTPSSLPGHT
EPGPLLIPFTFNFTITNLHYEENMQHPGSRKFNTTERVLQGLLKPLFKNTSVGPLYSGCRLTLLRPEKHEAATGVDTICTHRVDPIGPGLDRERLYWELSQLTNSITELGPYTLDRDSLYVNGFXXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
EPGPLLIPFTFNFTITNLHYEENMQHPGSRKFNTTERVLQGLLKPLFKNTSVGPLYSGCRLTLLRPEKHEAATGVDTICTHRVDPIGPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
APVPLLIPFTLNFTITNLHYEENMQHPGSRKFNTTERVLQGLLKPLFKSTSVGPLYSGCRLTLLRPEKNGAATGMDAICSHRLDPKSPGLDREQLYWQLSQMTNGIKELGPYTLDRNSLYVNGFTHQSSMTTTRTPDTSTMHLATSRTPASLSGPT
SAGPLLVPFTLNFTITNLQYEEDMHHPGSRKFNTTERVLQGLLGPLFKNSSVGPLYSGCRLISLRSEKDGAATGMDAVCLYHPNPKRPGLDREQLYWELSQLTHNITELGPYSLDRDSLYVNGFTHQSSVSTTSTPGTSTVDLRTSGTPSSLSSPTIM
TAVPLLVPFTLNFTITNLQYGEDMRHPGSRKFNATERVLQGLLSPIFKNSSVGPLYSGCRLTSLRPEKDGAATGMDAVCLYHPNPKRPGLDREQLYCELSQLTHNITELGPYSLDRDSLYVNGFNPRSSVPTTSTPGTSTVVHLATSGTPSSLPGHT
TAGPLLVPFTLNFTITNLQYEEDMHRPGSRKFNATERVLQGLLSPIFKNSSVGPLYSGCRLTLLRPKKDGAATGVDAICTHRLDPKSPGLDREQLYWELSQLTHSITELGPYTLDRVSLYVNGFNPRSSVPTTSIPGTSAVHLETSGTPASLPGHT
TASPLLVLFTINCTITNLQYEEDMRRTGSRKFNTMESVLQGLLKPLFKNTSVSSLYSGCRLTLLRPEKDGAATRVDAACTYRPDPKSPGLDREQLYWELSQLTNSVTELGPYTLDRDSLYVNGFTQRSSVPTTSIPGTSAVHLETSGTPASLPGHT
XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLRPLFKSTSVGPLYSGCRLTLLRPEKHGAATGVDTICTHRLDPLNPGLDREQLYWELSKLTRGIIELGPYLLDRGSLYVNGFTHRNFVPITSTPGTSTVHLGTSETPSSLPRPI
APVPLLIPFTLNFTITNLHYEENMQHPGSRKFNTTERVLQGLLRPLFKSTSVGPLYSGCRLTLLRPEKRGAATGVDTICTHRLDPLNPGLDREQLYWELSQLTHGITELGPYTLDRDSLYVDGFTHWSPIPTTSTPGTSIVNLGTSGIPPSLPETT
APGPLLVPFTLNFTITNLQYEVDMRHPGSRKFNTTERVLQGLLRPLFKNTSIGPLYSSCRLTLLRPEKDKAATRVDAICTHHPDPQSPGLNREQLYWELSNLTNGIQELGPYTLDRNSLYVNGFNPWSSVPTTSTPGTSTVHLATSGTPSSLPGHT
VPGPLLVPFTLNFTITNLQYEEAMRHPGSRKFNTTERVLQGLLKPLFRNSSLEYLYSGCRLASLRPEKDSSAMAVDAICTHRVDPIGPGLDRERLYWELSQLTNSITELGPYTLDRDSLYVNGFTHRSSVPTTSIPGTSAVHLETSGTPASLPGHT
XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLKPLFRNSSLEYLYSGCRLASLRPEKDSSAMAVDAICTHRPDPEDLGLDRERLYWELSNLTNGIQELGPYTLDRNSLYVNGFTHRSSMPTTSTPGTSTVDLGTSGTPSSLPGHT
TAGPLLVPFTLNFTITNLQYEEDMRHPGSRRFNTTERVLQGLLTPLFKNTSVGPLYSGCRLTLLRPEKHGAATGVDAICTLRLDPTGPGLDRERLYWELSQLTHGITELGPYTLDRHSLYVNGFTHQSSMTTTRTPDTSTMHLATSRTPASLSGPT
APVPLLIPFTLNFTITDLHYEENMQHPGSRKFNTTERVLQGLLKPLFKNTSVSSLYSGCRLTLLRPEKDGAATRVDAVCTHRPDPKSPGLDREQLYWELSQLTHSITELGPYTQDRDSLYVNGFTHRSSVPTTSTPGTSTVDLGTSGTPFSLPSPA
APGPLLVPFTLNFTITNLQYEEDMRHPGSRKFSTTERVLQGLLKPLFKNTSVSSLYSGCRLTLLRPEKDGAATKVDAICTYRPDPKSPGLDREQLYWELSKLTRGIIELGPYLLDRGSLYVNGFTHRTSVPTTSTPGTSTVDLG.SGTPSLPSSPT
TASPLLVLFTINFTITNQRYEENMHHPGSRKFNTTERVLQGLLRPVFKNTSVGPLYSGCRLTLLRPEKRGAATGVDTICTHRLDPLNPGLDREQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHQTSAPNTSTPGTSTVDLGTSGTPSSLPSPT
APGPLLVPFTLNFTITNLQYEEDMRHPGSRKFNTTERVLQTLLGPMFKNTSVGLLYSGCRLTLLRSEKDGAATGVDAICTHRLDPKSPGVDREQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHWIPVPTSSTPGTSTVDLG.SGTPSSLPSPT
XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQTLLGPMFKNTSVGPLYSGCRLTLLRSEKDGAATGVDAICTHRLDPKSPGVDREQLYWELSXLTXXIXELGPYXLDRXSLYVNGFTHWIPVPTSSTPGTSTVDLGTSGTPSSLPSPT
TAGPLLVPFTLNFTITNLKYEEDMHCPGSRKFNTTERVLQSLLGPMFKNTSVGPLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTNGIKELGPYTLDRNSLYVNGFTHQTSAPNTSTPGTSTVDLGTSGTPSSLPSPT
XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLXPXFKXTSVGXLYSGCRLTLLRSEKDGAATGVDAICTHRVDPKSPGVDREQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHRSSMPTTSTPGTSTVDVGTSGTPSSSPSPT
TAGPLLVPFTLNFTITNLKYEEDMRHPGSRKFNTTERVLQSLLGPMFKNTSVGPLYSGCRLTLLRSEKDGAATGVDAICTHRLDPKSPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
SAGPLLVPFTLNFTITNLQYEEDMHHPGSRKFNTTERVLQGLLGPMFKNTSVGLLYSGCRLASLRPEKDSSAMAVDAICTHRPDPEDLGLDRERLYWELSNLTNGIQELGPYTLDRNSLYVNGFTHRTSVPTTSTPGTSTVDLGTSGTPFSLPSPA
XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLKPLFRNSSLEYLYSGCRLASLRPEKDSSAMAVDAICTHRLDPKSPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
TAGPLLIPFTLNFTITNLQYGEDMGHPGSRKFNTTERVLQTLLGPMFKNTSVGLLYSGCRLTLLRSEKDGAATGVDAICTHRLDPKSPGLDREQLYWELSQLTHSITELGPYTQDRDSLYVNGFTHRSSVPTTSIPGTSAVHLETTGTPSSFPGHT
TAGPLLVLFTLNFTITNLKYEEDMHRPGSRKFNTTERVLQGLLGPIFKNTSVGPLYSGCRLTLLRPKKDGAATKVDAICTYRPDPKSPGLDREQLYWELSQLTHSITELGPYTQDRDSLYVNGFTHRSSVPTTSIPGTSAVHLETSGTPSSLPGHT
XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLRPVFKNTSVGPLYSGCRLTLLRPEKQEAATGVDTICTHRVDPIGPGLDRERLYWELSQLTNSITELGPYTLDRDSLYVNGFNPWSSVPTTSTPGTSTVHLATSGTPSSLPGHT
EPGPLLIPFTFNFTITNLRYEENMQHPGSRKFNTTERVLQGLLTPLFKNTSVGPLYSGCRLTLLRPEKHGAATGVDAICTLRLDPTGPGLDRERLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXXTSTPGTSXVXLXTSGTPASLPGHT
APVPLLIPFTLNFTITDLHYEENMQHPGSRKFNTTERVLQSLHGPMFKNTSVGPLYSGCRLTLLRSEKDGAATGVDAICTHRLDPKSPGLDREXLYWELSXLTNSITELGPYTLDRDSLYVNGFTHRSSMPTTSIPGTSAVHLETSGTPASLPGHT
TAGPLLVPFTLNFTITNLKYEEDMHCPGSRKFNTTERVLQGLLXPXFKXTSVGXLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTNSITELGPYTLDRDSLYVNGFTHRSSMPTTSIPGTSAVHLETSGTPASLPGHT
XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLKPLFKSTSVGXLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTNSITELGPYTLDRDSLYVNGFTHRSSMPTTSIPGTSAVHLETSGTPASLPGHT

TABLE 16 - continued

CA125 Repeat Domains
(SEQ ID NO: 146)

```
40  APGPLLVPFTLNFTITNLQYEEDMRHPGSRKFNTTERVLQGLLKPLFKSTSVGPLYSGCRLTLLRPEKRGAATGVDTICTHRLDPLNPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
    XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLXPXFKXTSVGXLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
    APVPLLIPFTLNFTITNLHYEENMQHPGSRKFNTTERVLQGLLGPMFKNTSVGLLYSGCRLTLLRPEKNGAATGMDAICSHRLDPKSPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFTHQNSVPTTSTPGTSTVYWATTGTPSSFPGHT
    XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLXPXFKXTSVGXLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
    EPGPLLIPFTFNFTITNLHYEENMQHPGSRKFNTTERVLQGLLTPLFKNTSVGPLYSGCRLTLLRPEKQEAATGVDTICTHRVDPIGPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFTHRSSVPTTSSPGTSTVHLATSGTPSSLPGHT
45  XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLKPLFKSTSVGPLYSGCRLTLLRPEKHGAATGVDAICTLRLDPTGPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFTHRTSVPTTSTPGTSTVHLATSGTPSSLPGHT
    APVPLLIPFTLNFTITNLHYEENMQHPGSRKFNTTERVLQGLLKPLFKSTSVGPLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
    XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLXPXFKXTSVGXLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFTHWSSGLTTSTPWTSTVDLGTSGTPSPVPSPT
    APVPLLIPFTLNFTITNLQYEEDMHRPGSRKFNTTERVLQGLLSPIFKNTSVGPLYSGCRLTSLRPEKDGAATGMDAVCLYHPNPKRPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFTHWSSGLTTSTPWTSTVDLGTSGTPSPVPSPT
    XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLXPXFKXTSVGXLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
    APVPLLIPFTLNFTITNLQYEEDMHRPGSRKFNTTERVLQGLLXPXFKXTSVGXLYSGCRLTLLRPEKQEAATGVDTICTHRVDPIGPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFTHRSPGLTTSTPWTSTVDLGTSGTPSPVPSPT
    XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNATERVLQGLLSPIFKNTSVGPLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
50  TAGPLLVPFTLNFTITNLQYEEDMHRPGSRKFNTTERVLQGLLXPXFKXTSVGXLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
    XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLTPLFRNTSVSSLYSGCRLTLLRPEKDGAATRVDAVCTHRPDPKSPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFTHWIPVPTSSTPGTSTVDLG.SGTPSSLPSPT
    TAGPLLVPFTLNFTITNLQYEEDMHRPGSRKFNTTERVLQGLLXPXFKXTSVGXLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
    XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLXPXFKXTSVGXLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFTHQTFAPNTSTPGTSTVDLGTSGTPSSLPSPT
    TAGPLLVPFTLNFTITNLQYGEDMGHPGSRKFNTTERVLQGLLGPIFKNTSVGPLYSGCRLTSLRSEKDGAATGVDAICIHHLDPKSPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFTHQTFAPNTSTPGTSTVDLGTSGTPSSLPSPT
    XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLXPXFKXTSVGXLYSGCRLTLLRXEKXXAATXVDXXCXXXXDPXXPGLDREXLYWELSXLTXXIXELGPYXLDRXSLYVNGFXXXXXXXXTSTPGTSXVXLXTSGTPXXXPXXT
55  SAGPLLVPFTLNFTITNLQYEEDMHHPGSRKFNTTERVLQGLLGPMFKNTSVGLLYSGCRLTLLRPEKNGAATRVDAVCTHRPDPKSPGLDREXLYWELSQLTHSITELGPYTLDRDSLYVNGFTQRSSVPTTSTPGTFTVQPETSETPSSLPGPT
    XXXPLLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERVLQGLLKPLFKSTSVGPLYSGCRLTLLRPEKDGVATRVDAICTHRPDPKIPGLDRQQLYWELSQLTHSITELGPYTLDRHSLYVNGFTHQSSMTTTRTPDTSTMHLATSRTPASLSGPT
    ATGPVLLPFTLNFTITNLQYEEDMHRPGSRKFNTTERVLQGLLMPLFKNTSVSSLYSGCRLTLLRPEKDGAATRVDAVCTHRPDPKSPGLDRERLYWKLSQLTHGITELGPYTLDRHSLYVNGFTQRSSVPTTSVPGTPTVDLGTSGTPVSKPGPS
    TASPLLVLFTINFTITNLRYEENMHHPGSRKFNTTERVLQGLLRPVFKNTSVGPLYSGCRLTLLRPEKDGTATGVDAICTHHPDPKSPRLDREQLYWELSQLTHNITELGHYALDNDSLFVNGFTHRSSVSTTSTPGTPTVYLGASKTPASIFGPS
60  AASPLLVLFTLNGTITNLRYEENMQHPGSRKFNTTERVLQGLLRPLFKNTSVGPLYSGSRLTLLRPEKDGEATGVDAICTHRPDPTGPGLDREQLYLELSQLTHSITELGPYTLDRDSLYVNGFTHRSSVPTTS.....
    AASHLLILFTLNFTITNLRYEENMW.PGSRKFNTTERVLQGLLRPLFKNTSVGPLYSGSRLTLLRPEKDGEATGVDAICTHRPDPTGPGLDREQLYLELSQLTHSITELGHYALDNDSLFVNGFTHRSSVPTTS.....
    TGVVSEEPFTLNFTINNLRYMADMGQPGSLKFNITDNVMKHLLSPLFQRSSLGARYTGCRVIALRSVKNGAETRVDLLCTYLQPLSGPGLPIKQVFHELSQQTHGITRLGPYSLDKDSLYLNGYNEPGLDEPPTTPKPATTFLPPLSEATT.....
    AMGYHLKTLTLNFTISNLQYSPDMGKGSATFNSTEGVLQHLLRPLFQKSSM.GPFYLGCQLISLRPEKDGAATGVDTTCTYHPDPVGPGLDIQQLYWELSQLTHGVTQLGFYVLDRDSLFINGYAPQNLSIRGEYQINFHIVNWNLSNPDPTSSEY
```

TABLE 17

Carboxy Terminal Nucleotide Sequence (SEQ ID NO: 147)	
1	GCCATGGGGT ACCACCTGAA GACCCTCACA CTCAACTTCA CCATCTCCAA
51	TCTCCAGTAT TCACCAGATA TGGGCAAGGG CTCAGCTACA TTCAACTCCA
101	CCGAGGGGGT CCTTCAGCAC CTGCTCAGAC CCTTGTTCCA GAAGAGCAGC
151	ATGGGCCCCT TCTACTTGGG TTGCCAACTG ATCTCCCTCA GGCCTGAGAA
201	GGATGGGGCA GCCACTGGTG TGGACACCAC CTGCACCTAC CACCCTGACC
251	CTGTGGGCCC CGGGCTGGAC ATACAGCAGC TTTACTGGGA GCTGAGTCAG
301	CTGACCCATG GTGTCACCCA ACTGGGCTTC TATGTCCTGG ACAGGGATAG
351	CCTCTTCATC AATGGCTATG CACCCCAGAA TTTATCAATC CGGGGCGAGT
401	ACCAGATAAA TTTCCACATT GTCAACTGGA ACCTCAGTAA TCCAGACCCC
451	ACATCCTCAG AGTACATCAC CCTGCTGAGG GACATCCAGG ACAAGGTCAC
501	CACACTCTAC AAAGGCAGTC AACTACATGA CACATTCCGC TTCTGCCTGG
551	TCACCAACTT GACGATGGAC TCCGTGTTGG TCACTGTCAA GGCATTGTTC
601	TCCTCCAATT TGGACCCCAG CCTGGTGGAG CAAGTCTTTC TAGATAAGAC
651	CCTGAATGCC TCATTCCATT GGCTGGGCTC CACCTACCAG TTGGTGGACA
701	TCCATGTGAC AGAAATGGAG TCATCAGTTT ATCAACCAAC AAGCAGCTCC
751	AGCACCCAGC ACTTCTACCT GAATTTCACC ATCACCAACC TACCATATTC
801	CCAGGACAAA GCCCAGCCAG GCACCACCAA TTACCAGAGG AACAAAAGGA
851	ATATTGAGGA TGCGCTCAAC CAACTCTTCC GAAACAGCAG CATCAAGAGT
901	TATTTTTCTG ACTGTCAAGT TTCAACATTC AGGTCTGTCC CCAACAGGCA

TABLE 17-continued

Carboxy Terminal Nucleotide Sequence (SEQ ID NO: 147)	
951	CCACACCGGG GTGGACTCCC TGTGTAACTT CTCGCCACTG GCTCGGAGAG *
1001	TAGACAGAGT TGCCATCTAT GAGGAATTTC TGCGGATGAC CCGGAATGGT
1051	ACCCAGCTGC AGAACTTCAC CCTGGACAGG AGCAGTGTCC TTGTGGATGG
1101	GTATTCTCCC AACAGAAATG AGCCCTTAAC TGGGAATTCT GACCTTCCCT
1151	TCTGGGCTGT CATCCTCATC GGCTTGGCAG GACTCCTGGG ACTCATCACA
1201	TGCCTGATCT GCGGTGTCCT GGTGACCACC CGCCGGCGGA AGAAGGAAGG
1251	AGAATACAAC GTCCAGCAAC AGTGCCCAGG CTACTACCAG TCACACCTAG
1301	ACCTGGAGGA TCTGCAATGA CTGGAACTTG CCGGTGCCTG GGGTGCCTTT
1351	CCCCCAGCCA GGGTCCAAAG AAGCTTGGCT GGGGCAGAAA TAAACCATAT
1401	TGGTCGGAAA AAAAAAAAA AA

TABLE 18

Carboxy Terminal Amino Acid Sequence
(SEQ ID NO: 148)

```
  1  AMGYHLKTLT LNFTISNLQY SPDMGKGSAT FNSTEGVLQH LLRPLFQKSS
 51  MGPFYLGCQL ISLRPEKDGA ATGVDTTCTY HPDPVGPGLD IQQLYWELSQ
101  LTHGVTQLGF YVLDRDSLFI NGYAPQNLSI RGEYQINFHI VNWNLSNPDP
                *
151  TSSEYITLLR DIQDKVTTLY KGSQLHDTFR FCLVTNLTMD SVLVTVKALF
201  SSNLDPSLVE QVFLDKTLNA SFHWLGSTYQ LVDIHVTEME SSVYQPTSSS
251  STQHFYLNFT ITNLPYSQDK AQPGTTNYQR NKRNIEDALN QLFRNSSIKS
301  YFSDCQVSTF RSVPNRHHTG VDSLCNFSPL ARRVDRVAIY EEFLRMTRNG
351  TQLQNFTLDR SSVLVDGYSP NRNEPLTGNS DLPFWAVILI GLAGLLGLIT
401  CLICGVLVTT RRRKKEGEYN VQQQCPGYYQ SHLDLEDLQ
```

TABLE 19A

Serine/Threonine O-glycosylation Pattern Predicted for the Amino Terminal End of the CA125 Molecule (SEQ ID NO: 149)

SEQ ID NO: 149 Length: 1799

RTDGIMEHITKIPNEAAHRGTIRPVKGPQTSTSPASPKGLHTGGTKRMETTTTALKTTTTALKTTSRATLTTSVYTPTLG 80
TLTPLNASRQMASTILTEMMITTPYVFPDVPETTSSLATSLGAETSTALPRTTPSVLNRESETTASLVSRSGAERSPVIQ 160
TLDVSSSEPDTTASWVIHPAETIPTVSKTTPNFFHSELDTVSSTATSHGADVSSAIPTNISPSELDALTPLVTISGTDTS 240
TTFPTLTKSPHETETRTTWLTHPAETSSTIPRTIPNFSHHESDATPSIATSPGAETSSAIPIMTVSPGAEDLVTSQVTSS 320
GTDRNMTIPTLTLSPGEPKTIASLVTHPEAQTSSAIPTSTISPAVSRLVTSMVTSLAAKTSTTNRALTNSPGEPATTVSL 400
VTHPAQTSPTVPWTTSIFFHSKSDTTPSMTTSHGAESSSAVPTPTVSTEVPGVVTPLVTSSRAVISTTIPILTLSPGEPE 480
TTPSMATSHGEEASSAIPTPTVSPGVPGVVTSLVTSSRAVTSTTIPILTFSLGEPETTPSMATSHGTEAGSAVPTVLPEV 560
PGMVTSLVASSRAVTSTTLPTLTLSPGEPETTPSMATSHGAEASSTVPTVSPEVPGVVTSLVTSSSGVNSTSIPTLILSP 640
GELETTPSMATSHGAEASSAVPTPTVSPGVSGVVTPLVTSSRAVTSTTIPILTLSSSEPETTPSMATSHGVEASSAVLTV 720
SPEVPGMVTSLVTSSRAVTSTTIPTLTISSDEPETTTSLVTHSEAKMISAIPTLAVSPTVQGLVTSLVTSSGSETSAFSN 800
LTVASSQPETIDSWVAHPGTEASSVVPTLTVSTGEPFTNISLVTHPAESSSTLPRTTSRFSHSELDTMPSTVTSPEAESS 880
SAISTTISPGIPGVLTSLVTSSGRDISATFPTVPESPHESEATASWVTHPAVTSTTVPRTTPNYSHSEPDTTPSIATSPG 960
AEATSDFPTITVSPDVPDMVTSQVTSSGTDTSITIPTLTLSSGEPETTTSFITYSETHTSSAIPTLPVSPGASKMLTSLV 1040
ISSGTDSTTTFPTLTETPYEPETTAIQLIHPAETNTMVPRTTPKFSHSKSDTTLPVAITSPGPEASSAVSTTTISPDMSD 1120
LVTSLVPSSGTDTSTTFPTLSETPYEPETTATWLTHPAETSTTVSGTIPNFSHRGSDTAPSMVTSPGVDTRSGVPTTTIP 1200
PSIPGVVTSQVTSSATDTSTAIPTLTPSPGEPETTASSATHPGTQTGFTVPIRTVPSSEPDTMASWVTHPPQTSTPVSRT 1280
TSSFSHSSPDATPVMATSPRTEASSAVLTTISPGAPEMVTSQITSSGAATSTTVPTLTHSPGMPETTALLSTHPRTETSK 1360
TFPASTVFPQVSETTASLTIRPGAETSTALPTQTTSSLFTLLVTGTSRVDLSPTASPGVSAKTAPLSTHPGTETSTMIPT 1440
STLSLGLLETTGLLATSSSAETSTSTLTLTVSPAVSGLSSASITTDKPQTVTSWNTETSPSVTSVGPPEFSRTVTGTTMT 1520
LIPSEMPTPPKTSHGEGVSPTTILRTTMVEATNLATTGSSPTVAKTTTTFNTLAGSLFTPLTTPGMSTLASESVTSRTSY 1600
NHRSWISTTSSYNRRYWTPATSTPVTSTFSPGISTSSIPSSTAATVPFMVPFTLNFTITNLQYEEDMRHPGSRKFNATER 1680
ELQGLLKPLFRNSSLEYLYSGCRLASLRPEKDSSAMAVDAICTHRPDPEDLGLDRERLYWELSNLTNGIQELGPYTLDRN 1760
SLYVNGFTHRSSMPTTSTPGTSTVDVGTSGTPSSSPSPT

TABLE 19B

....................T........TSTS................TTT....TTTT...TT.....TT...T.... 80
....ST....TT........................... 160
.....S....T.............T.S................T........S.........S...........S.T..S 240
T...T.T..................TSS....T........S..T.S..TS......S.....T.........T...TS. 320
...........T.S.....T..S........TSS...TST.............T......STT....T.S.....TT.S. 400
.T....TS.T...T........S..T...TTS....SSS...T.T.ST..................T.....T.S..... 480
TT.S..T......SS...T.T.S.............S.....T..............T.S..TS......S...T..... 560
.................T....T.S.....TT.S..TS.....SST..T.S...........TS.S....T......... 640
.....T.S..T......SS...T.T.S...S.........S.....T.....T.SSS....T.S..TS......S..... 720
S.............S....STT..T.T.SS.....TT...........S...................T........... 800
....S.................SS.....T.............T....SSS....T.............ST.T......S 880
S...TT.S..................S....T.......S..T....T....TSTT...TT...S.S....T.S..TS.. 960
...TS.....T.........T...TS..........T.T.SS.....T....T.....T.S...T............... 1040
.S..T.STTT..T.T.T..................T....TT.......S.........S.....SS....TT....... 1120
.......S..T..STT..T.S.T.....TT....T.....ST.....................TS......S....TT.. 1200
.S.....T...TS..T.TST...T.T.S.....TT.SS.T........T....T..SS...T..S..T....TST..S.T 1280
TSS.S.SS...T....TS..T..SS....T.S.......T...TS....TSTT....T.S..........ST...T..S. 1360
....ST.....S.TT...T.......ST...T.TT.S................T.S...S......ST...T..ST...T 1440
ST.............T..S..TSTS....T.....S..S..S..T....T.TS..T..S.S.TS......S........T 1520

TABLE 19B-continued

Serine/Threonine O-glycosylation Pattern Predicted for the Amino Terminal End of the CA125 Molecule	
.....T...ST..S.....	1600
.....TT.SS.T.....	1680
...S...T....S.....T.....TST..TST.S...STSS..SST.....	1760
.....	
.....	
.....TTST...ST....TS.T.SSS.S.T	

TABLE 20

Nucleotide and Amino Acid Sequences of Recombinant CA125 Repeat Showing Peptides (Underlined 1-4) which are Antigenically Matched for Immune Stimulation of Patients with the HLA-2 Histocompatibility Subtype

CA 125 Recombinant Nucleotide and Amino Acid Sequences (SEQ ID NO: 151 and SEQ ID NO: 152, respectively)
CA 125 Recombinant Nucleotide (Anti-Sense Strand) Sequence (SEQ ID NO: 153)
Peptide 1 (SEQ ID NO: 154); Peptide 2 (SEQ ID NO: 155);
Peptide 3 (SEQ ID NO: 156) and Peptide 4 (SEQ ID NO: 157)

```
      ATGAGAGGATCGCATCACCATCACCATCACGGATCCATGGGCCACACAGAGCCTGGCCCT
    1 ---------+---------+---------+---------+---------+---------+ 60
      TACTCTCCTAGCGTAGTGGTAGTGGTAGTGCCTAGGTACCCGGTGTGTCTCGGACCGGGA

      M  R  G  S  H  H  H  H  H  H  G  S  M  G  H  T  E  P  G  P   -
                                                      ↑
      CTCCTGATACCATTCACTTTCAACTTTACCATCACCAACCTGCATTATGAGGAAAACATG
   61 ---------+---------+---------+---------+---------+---------+ 120
      GAGGACTATGGTAAGTGAAAGTTGAAATGGTAGTGGTTGGACGTAATACTCCTTTTGTAC

      L  L  I  P  F  T  F  N  F  T  I  T  N  L  H  Y  E  E  N  M   -

      CAACACCCTGGTTCCAGGAAGTTCAACACCACGGAGAGGGTTCTGCAGGGTCTGCTCAAG
  121 ---------+---------+---------+---------+---------+---------+ 180
      GTTGTGGGACCAAGGTCCTTCAAGTTGTGGTGCCTCTCCCAAGACGTCCCAGACGAGTTC
                                                    3
      Q  H  P  G  S  R  K  F  N  T  T  E  R  V  L  Q  G  L  L  K   -

      CCCTTGTTCAAGAACACCAGTGTTGGCCCTCTGTACTCTGGCTGCAGACTGACCTTGCTC
  181 ---------+---------+---------+---------+---------+---------+ 240
      GGGAACAAGTTCTTGTGGTCACAACCGGGAGACATGAGACCGACGTCTGACTGGAACGAG

      P  L  F  K  N  T  S  V  G  P  L  Y  S  G  C  R  L  T  L  L   -

      AGACCTGAGAAGCATGAGGCAGCCACTGGAGTGGACACCATCTGTACCCACCGCGTTGAT
  241 ---------+---------+---------+---------+---------+---------+ 300
      TCTGGACTCTTCGTACTCCGTCGGTGACCTCACCTGTGGTAGACATGGGTGGCGCAACTA

      R  P  E  K  H  E  A  A  T  G  V  D  T  I  C  T  H  R  V  D   -

      CCCATCGGACCTGGACTGGACAGAGAGCGGCTATACTGGGAGCTGAGCCAGCTGACCAAC
  301 ---------+---------+---------+---------+---------+---------+ 360
      GGGTAGCCTGGACCTGACCTGTCTCTCGCCGATATGACCCTCGACTCGGTCGACTGGTTG
                                1                       4
      P  I  G  P  G  L  D  R  E  R  L  Y  W  E  L  S  Q  L  T  N   -

      AGCATCACAGAGCTGGGACCCTACACCCTGGACAGGGACAGTCTCTATGTCAATGGCTTC
  361 ---------+---------+---------+---------+---------+---------+ 420
      TCGTAGTGTCTCGACCCTGGGATGTGGGACCTGTCCCTGTCAGAGATACAGTTACCGAAG
```

TABLE 20 (continued)

Nucleotide and Amino Acid Sequences of Recombinant CA125 Repeat Showing Peptides (Underlined 1-4) which are Antigenically Matched for Immune Stimulation of Patients with the HLA-2 Histocompatibility Subtype

CA 125 Recombinant Nucleotide and Amino Acid Sequences (SEQ ID NO: 151 and SEQ ID NO: 152, respectively)
CA 125 Recombinant Nucleotide (Anti-Sense Strand) Sequence (SEQ ID NO: 153)
Peptide 1 (SEQ ID NO: 154); Peptide 2 (SEQ ID NO: 155); Peptide 3 (SEQ ID NO: 156) and Peptide 4 (SEQ ID NO: 157)

```
                                2
      S  I  T  E  L  G  P  Y  T  L  D  R  D  S  L  Y  V  N  G  F  -
      ‾‾‾‾‾‾‾‾‾‾‾‾‾‾‾‾        ‾  ‾  ‾  ‾  ‾  ‾  ‾  ‾  ‾
      AACCCTCGGAGCTCTGTGCCAACCACCAGCACTCCTGGGACCTCCACAGTGCACCTGGCA
 421  ---------+---------+---------+---------+---------+---------+ 480
      TTGGGAGCCTCGAGACACGGTTGGTGGTCGTGAGGACCCTGGAGGTGTCACGTGGACCGT

      N  P  R  S  S  V  P  T  T  S  T  P  G  T  S  T  V  H  L  A  -

      ACCTCTGGGACTCCATCCTCCCTGCCT
 481  ---------+---------+------- 507
      TGGAGACCCTGAGGTAGGAGGGACGGA

      T  S  G  T  P  S  S  L  P  -
```

(SEQ ID NO: 154)
Peptide 1 R L Y W E L S Q L

(SEQ ID NO: 155)
Peptide 2 T L D R D S L Y V

(SEQ ID NO: 156)
Peptide 3 V L Q G L L K P L

(SEQ ID NO: 157)
Peptide 4 Q L T N S I T E L

TABLE 21

CA125 Protein Sequence

(SEQ ID NO: 162)

```
   1 MEHITKIPNE AAHRGTIRPV KGPQTSTSPA SPKGLHTGGT KRMETTTTAL
  51 KTTTTALKTT SRATLTTSVY TPTLGTLTPL NASRQMASTI LTEMMITTPY
 101 VFPDVPETTS SLATSLGAET STALPRTTPS VLNRESETTA SLVSRSGAER
 151 SPVIQTLDVS SSEPDTTASW VIHPAETIPT VSKTTPNFFH SELDTVSSTA
 201 TSHGADVSSA IPTNISPSEL DALTPLVTIS GTDTSTTFPT LTKSPHETET
 251 RTTWLTHPAE TSSTIPRTIP NFSHHESDAT PSIATSPGAE TSSAIPIMTV
 301 SPGAEDLVTS QVTSSGTDRN MTIPTLTLSP GEPKTIASLV THPEAQTSSA
 351 IPTSTISPAV SRLVTSMVTS LAAKTSTTNR ALTNSPGEPA TTVSLVTHPA
 401 QTSPTVPWTT SIFFHSKSDT TPSMTTSHGA ESSSAVPTPT VSTEVPGVVT
 451 PLVTSSRAVI STTIPILTLS PGEPETTPSM ATSHGEEASS AIPTPTVSPG
 501 VPGVVTSLVT SSRAVTSTTI PILTFSLGEP ETTPSMATSH GTEAGSAVPT
 551 VLPEVPGMVT SLVASSRAVT STTLPTLTLS PGEPETTPSM ATSHGAEASS
 601 TVPTVSPEVP GVVTSLVTSS SGVNSTSIPT LILSPGELET TPSMATSHGA
 651 EASSAVPTPT VSPGVSGVVT PLVTSSRAVT STTIPILTLS SSEPETTPSM
 701 ATSHGVEASS AVLTVSPEVP GMVTSLVTSS RAVTSTTIPT LTISSDEPET
 751 TTSLVTHSEA KMISAIPTLA VSPTVQGLVT SLVTSSGSET SAFSNLTVAS
 801 SQPETIDSWV AHPGTEASSV VPTLTVSTGE PFTNISLVTH PAESSSTLPR
 851 TTSRFSHSEL DTMPSTVTSP EAESSSAIST TISPGIPGVL TSLVTSSGRD
 901 ISATFPTVPE SPHESEATAS WVTHPAVTST TVPRTTPNYS HSEPDTTPSI
 951 ATSPGAEATS DFPTITVSPD VPDMVTSQVT SSGTDTSITI PTLTLSSGEP
1001 ETTTSFITYS ETHTSSAIPT LPVSPGASKM LTSLVISSGT DSTTTFPTLT
1051 ETPYEPETTA IQLIHPAETN TMVPRTTPKF SHSKSDTTLP VAITSPGPEA
1101 SSAVSTTTIS PDMSDLVTSL VPSSGTDTST TFPTLSETPY EPETTATWLT
1151 HPAETSTTVS GTIPNFSHRG SDTAPSMVTS PGVDTRSGVP TTTIPPSIPG
1201 VVTSQVTSSA TDTSTAIPTL TPSPGEPETT ASSATHPGTQ TGFTVPIRTV
1251 PSSEPDTMAS WVTHPPQTST PVSRTTSSFS HSSPDATPVM ATSPRTEASS
1301 AVLTTISPGA PEMVTSQITS SGAATSTTVP TLTHSPGMPE TTALLSTHPR
1351 TETSKTFPAS TVFPQVSETT ASLTIRPGAE TSTALPTQTT SSLFTLLVTG
1401 TSRVDLSPTA SPGVSAKTAP LSTHPGTETS TMIPTSTLSL GLLETTGLLA
1451 TSSSAETSTS TLTLTVSPAV SGLSSASITT DKPQTVTSWN TETSPSVTSV
1501 GPPEFSRTVT GTTMTLIPSE MPTPPKTSHG EGVSPTTILR TTMVEATNLA
1551 TTGSSPTVAK TTTTFNTLAG SLFTPLTTPG MSTLASESVT SRTSYNHRSW
1601 ISTTSSYNRR YWTPATSTPV TSTFSPGIST SSIPSSTA
```

Amino Terminal Domain

TABLE 21 - continued

CA125 Protein Sequence

(SEQ ID NO: 162)

```
                                                                AT VPFMVPFTLN
1651  FTITNLQYEE DMRHPGSRKF NATERELQGL LKPLFRNSSL EYLYSGCRLA
1701  SLRPEKDSSA MAVDAICTHR PDPEDLGLDR ERLYWELSNL TNGIQELGPY
1751  TLDRNSLYVN GFTHRSSMPT TSTPGTSTVD VGTSGTPSSS PSPTAAGPLL
1801  MPFTLNFTIT NLQYEEDMRR TGSRKFNTME SVLQGLLKPL FKNTSVGPLY
1851  SGCRLTLLRP EKDGAATGVD AICTHRLDPK SPGLNREQLY WELSKLTNDI
1901  EELGPYTLDR NSLYVNGFTH QSSVSTTSTP GTSTVDLRTS GTPSSLSSPT
1951  IMAAGPLLVP FTLNFTITNL QYGEDMGHPG SRKFNTTERV LQGLLGPIFK
2001  NTSVGPLYSG CRLTSLRSEK DGAATGVDAI CIHHLDPKSP GLNRERLYWE
2051  LSQLTNGIKE LGPYTLDRNS LYVNGFTHRT SVPTSSTPGT STVDLGTSGT
2101  PFSLPSPATA GPLLVLFTLN FTITNLKYEE DMHRPGSRKF NTTERVLQTL
2151  LGPMFKNTSV GLLYSGCRLT LLRSEKDGAA TGVDAICTHR LDPKSPGLDR
2201  EQLYWELSQL TNGIKELGPY TLDRNSLYVN GFTHWIPVPT SSTPGTSTVD
2251  LGSGTPSSLP SPTAAGPLLV PFTLNFTITN LQYEEDMHHP GSRKFNTTER
2301  VLQGLLGPMF KNTSVGLLYS GCRLTLLRSE KDGAATGVDA ICTHRLDPKS
2351  PGVDREQLYW ELSQLTNGIK ELGPYTLDRN SLYVNGFTHQ TSAPNTSTPG
2401  TSTVDLGTSG TPSSLPSPTS AGPLLVPFTL NFTITNLQYE EDMRHPGSRK
2451  FNTTERVLQG LLKPLFKSTS VGPLYSGCRL TLLRSEKDGA ATGVDAICTH
2501  RLDPKSPGVD REQLYWELSQ LTNGIKELGP YTLDRNSLYV NGFTHQTSAP
2551  NTSTPGTSTV DLGTSGTPSS LPSPTSAGPL LVPFTLNFTI TNLQYEEDMH
2601  HPGSRKFNTT ERVLQGLLGP MFKNTSVGLL YSGCRLTLLR PEKNGAATGM
2651  DAICSHRLDP KSPGLNREQL YWELSQLTHG IKELGPYTLD RNSLYVNGFT
2701  HRSSVAPTST PGTSTVDLGT SGTPSSLPSP TTAVPLLVPF TLNFTITNLQ
2751  YGEDMRHPGS RKFNTTERVL QGLLGPLFKN SSVGPLYSGC RLISLRSEKD
2801  GAATGVDAIC THHLNPQSPG LDREQLYWQL SQMTNGIKEL GPYTLDRNSL
2851  YVNGFTHRSS GLTTSTPWTS TVDLGTSGTP SPVPSPTTAG PLLVPFTLNF
2901  TITNLQYEED MHRPGSRKFN ATERVLQGLL SPIFKNSSVG PLYSGCRLTS
2951  LRPEKDGAAT GMDAVCLYHP NPKRPGLDRE QLYWELSQLT HNITELGPYS
3001  LDRDSLYVNG FTHQNSVPTT STPGTSTVYW ATTGTPSSFP GHTEPGPLLI
3051  PFTFNFTITN LHYEENMQHP GSRKFNTTER VLQGLLKPLF KNTSVGPLYS
3101  GCRLTSLRPE KDGAATGMDA VCLYHPNPKR PGLDREQLYC ELSQLTHNIT
3151  ELGPYSLDRD SLYVNGFTHQ NSVPTTSTPG TSTVYWATTG TPSSFPGHTE
3201  PGPLLIPFTF NFTITNLHYE ENMQHPGSRK FNTTERVLQG LLKPLFKNTS
3251  VGPLYSGCRL TLLRPEKHEA ATGVDTICTH RVDPIGPGLD RERLYWELSQ
3301  LTNSITELGP YTLDRDSLYV NGFNPRSSVP TTSTPGTSTV HLATSGTPSS
3351  LPGHTAPVPL LIPFTLNFTI TNLHYEENMQ HPGSRKFNTT ERVLQGLLKP
3401  LFKNTSVGPL YSGCRLTLLR PEKHEAATGV DTICTHRVDP IGPGLDREXL
3451  YWELSXLTXX IXELGPYXLD RXSLYVNGFX XXXXXXXTST PGTSXVXLXT
3501  SGTPXXXPXX TSAGPLLVPF TLNFTITNLQ YEEDMHHPGS RKFNTTERVL
3551  QGLLGPMFKN TSVGLLYSGC RLTLLRPEKN GAATGMDAIC SHRLDPKSPG
3601  LDREQLYWEL SQLTHGIKEL GPYTLDRNSL YVNGFTHRSS VAPTSTPGTS
3651  TVDLGTSGTP SSLPSPTTAV PLLVPFTLNF TITNLQYGED MRHPGSRKFN
3701  TTERVLQGLL GPLFKNSSVG PLYSGCRLIS LRSEKDGAAT GVDAICTHHL
3751  NPQSPGLDRE QLYWQLSQMT NGIKELGPYT LDRNSLYVNG FTHRSSGLTT
3801  STPWTSTVDL GTSGTPSPVP SPTTAGPLLV PFTLNFTITN LQYEEDMHRP
3851  GSRKFNATER VLQGLLSPIF KNSSVGPLYS GCRLTSLRPE KDGAATGMDA
3901  VCLYHPNPKR PGLDREQLYW ELSQLTHNIT ELGPYSLDRD SLYVNGFTHQ
3951  SSMTTTRTPD TSTMHLATSR TPASLSGPTT ASPLLVLFTI NCTITNLQYE
```

Repeat Domain

TABLE 21 - continued

CA125 Protein Sequence

(SEQ ID NO: 162)

```
4001  EDMRRTGSRK FNTMESVLQG LLKPLFKNTS VGPLYSGCRL TLLRPKKDGA
4051  ATGVDAICTH RLDPKSPGLN REQLYWELSK LTNDIEELGP YTLDRNSLYV
4101  NGFTHQSSVS TTSTPGTSTV DLRTSGTPSS LSSPTIMXXX PLLXPFTLNF
4151  TITNLXYEEX MXXPGSRKFN TTERVLQGLL RPLFKNTSVS SLYSGCRLTL
4201  LRPEKDGAAT RVDAACTYRP DPKSPGLDRE QLYWELSQLT HSITELGPYT
4251  LDRVSLYVNG FNPRSSVPTT STPGTSTVHL ATSGTPSSLP GHTXX XPLL
4301  XPFTLNFTIT NLXYEEXMXX PGSRKFNTTE RVLQGLLKPL FRNSSLEYLY
4351  SGCRLASLRP EKDSSAMAVD AICTHRPDPE DLGLDRERLY WELSNLTNGI
4401  QELGPYTLDR NSLYVNGFTH RSSFLTTSTP WTSTVDLGTS GTPSPVPSPT
4451  TAGPLLVPFT LNFTITNLQY EEDMHRPGSR RFNTTERVLQ GLLTPLFKNT
4501  SVGPLYSGCR LTLLRPEKQE AATGVDTICT HRVDPIGPGL DRERLYWELS
4551  QLTNSITELG PYTLDRDSLY VNGFNPWSSV PTTSTPGTST VHLATSGTPS
4601  SLPGHTAPVP LLIPFTLNFT ITDLHYEENM QHPGSRKFNT TERVLQGLLK
4651  PLFKSTSVGP LYSGCRLTLL RPEKHGAATG VDAICTLRLD PTGPGLDRER
4701  LYWELSQLTN SVTELGPYTL DRDSLYVNGF THRSSVPTTS IPGTSAVHLE
4751  TSGTPASLPG HTAPGPLLVP FTLNFTITNL QYEEDMRHPG SRKFSTTERV
4801  LQGLLKPLFK NTSVSSLYSG CRLTLLRPEK DGAATRVDAV CTHRPDPKSP
4851  GLDRERLYWK LSQLTHGITE LGPYTLDRHS LYVNGFTHQS SMTTTRTPDT
4901  STMHLATSRT PASLSGPTTA SPLLVLFTIN FTITNQRYEE NMHHPGSRKF
4951  NTTERVLQGL LRPVFKNTSV GPLYSGCRLT LLRPKKDGAA TKVDAICTYR
5001  PDPKSPGLDR EQLYWELSQL THSITELGPY TQDRDSLYVN GFTHRSSVPT
5051  TSIPGTSAVH LETSGTPASL PGHTAPGPLL VPFTLNFTIT NLQYEEDMRH
5101  PGSRKFNTTE RVLQGLLKPL FKSTSVGPLY SGCRLTLLRP EKRGAATGVD
5151  TICTHRLDPL NPGLDREQLY WELSKLTRGI IELGPYLLDR GSLYVNGFTH
5201  RTSVPTTSTP GTSTVDLGTS GTPFSLPSPA XXXPLLXPFT LNFTITNLXY
5201  EEXMXXPGSR KFNTTERVLQ TLLGPMFKNT SVGLLYSGCR LTLLRSEKDG
5251  AATGVDAICT HRLDPKSPGV DREQLYWELS QLTNGIKELG PYTLDRNSLY
5301  VNGFTHWIPV PTSSTPGTST VDLGSGTPSL PSSPTTAGPL LVPFTLNFTI
5351  TNLKYEEDMH CPGSRKFNTT ERVLQSLLGP MFKNTSVGPL YSGCRLTLLR
5401  SEKDGAATGV DAICTHRLDP KSPGVDREQL YWELSQLTNG IKELGPYTLD
5451  RNSLYVNGFT HQTSAPNTST PGTSTVDLGT SGTPSSLPSP TXXXPLLXPF
5501  TLNFTITNLX YEEXMXXPGS RKFNTTERVL QGLLXPXFKX TSVGXLYSGC
5551  RLTLLRXEKX XAATXVDXXC XXXXDPXXPG LDREXLYWEL SXLTXXIXEL
5601  GPYXLDRXSL YVNGFTHWIP VPTSSTPGTS TVDLGSGTPS SLPSPTTAGP
5651  LLVPFTLNFT ITNLKYEEDM HCPGSRKFNT TERVLQSLLG PMFKNTSVGP
5701  LYSGCRLTSL RSEKDGAATG VDAICTHRVD PKSPGVDREQ LYWELSQLTN
5751  GIKELGPYTL DRNSLYVNGF THQTSAPNTS TPGTSTVDLG TSGTPSSLPS
```

Repeat Domain

TABLE 21 - continued

CA125 Protein Sequence

(SEQ ID NO: 162)

```
5801  PTSAGPLLVP FTLNFTITNL QYEEDMHHPG SRKFNTTERV LQGLLGPMFK
5851  NTSVGLLYSG CRLTLLRPEK NGAATGMDAI CTHRLDPKSP GLDREXLYWE
5901  LSXLTXXIXE LGPYXLDRXS LYVNGFXXXX XXXXTSTPGT SXVXLXTSGT
5951  PXXXPXXTXX XPLLXPFTLN FTITNLXYEE XMXXPGSRKF NTTERVLQGL
6001  LKPLFRNSSL EYLYSGCRLA SLRPEKDSSA MAVDAICTHR PDPEDLGLDR
6051  ERLYWELSNL TNGIQELGPY TLDRNSLYVN GFTHRSSMPT TSTPGTSTVD
6101  VGTSGTPSSS PSPTTAGPLL IPFTLNFTIT NLQYGEDMGH PGSRKFNTTE
6151  RVLQGLLGPI FKNTSVGPLY SGCRLTSLRS EKDGAATGVD AICIHHLDPK
6201  SPGLNRERLY WELSQLTNGI KELGPYTLDR NSLYVNGFTH RTSVPTTSTP
6251  GTSTVDLGTS GTPFSLPSPA TAGPLLVLFT LNFTITNLKY EEDMHRPGSR
6301  KFNTTERVLQ TLLGPMFKNT SVGLLYSGCR LTLLRSEKDG AATGVDAICT
6351  HRLDPKSPGL DREXLYWELS XLTXXIXELG PYXLDRXSLY VNGFXXXXXX
6401  XXTSTPGTSX VXLXTSGTPX XXPXXTXXXP LLXPFTLNFT ITNLXYEEXM
6451  XXPGSRKFNT TERVLQGLLR PVFKNTSVGP LYSGCRLTLL RPKKDGAATK
6501  VDAICTYRPD PKSPGLDREQ LYWELSQLTH SITELGPYTQ DRDSLYVNGF
6551  THRSSVPTTS IPGTSAVHLE TTGTPSSFPG HTEPGPLLIP FTFNFTITNL
6601  RYEENMQHPG SRKFNTTERV LQGLLTPLFK NTSVGPLYSG CRLTLLRPEK
6651  QEAATGVDTI CTHRVDPIGP GLDRERLYWE LSQLTNSITE LGPYTLDRDS
6701  LYVDGFNPWS SVPTTSTPGT STVHLATSGT PSPLPGHTAP VPLLIPFTLN
6751  FTITDLHYEE NMQHPGSRKF NTTERVLQGL LKPLFKSTSV GPLYSGCRLT
6801  LLRPEKHGAA TGVDAICTLR LDPTGPGLDR ERLYWELSQL TNSITELGPY
6851  TLDRDSLYVN GFNPWSSVPT TSTPGTSTVH LATSGTPSSL PGHTTAGPLL
6901  VPFTLNFTIT NLKYEEDMHC PGSRKFNTTE RVLQSLHGPM FKNTSVGPLY
6951  SGCRLTLLRS EKDGAATGVD AICTHRLDPK SPGLDREXLY WELSXLTXXI
7001  XELGPYXLDR XSLYVNGFXX XXXXXXTSTP GTSXVXLXTS GTPXXXPXXT
7051  XXXPLLXPFT LNFTITNLXY EEXMXXPGSR KFNTTERVLQ GLLXPXFKXT
7101  SVGXLYSGCR LTLLRXEKXX AATXVDXXCX XXXDPXXPGL DREXLYWELS
7151  XLTNSITELG PYTLDRDSLY VNGFTHRSSM PTTSIPGTSA VHLETSGTPA
7201  SLPGHTAPGP LLVPFTLNFT ITNLQYEEDM RHPGSRKFNT TERVLQGLLK
7251  PLFKSTSVGP LYSGCRLTLL RPEKRGAATG VDTICTHRLD PLNPGLDREX
7301  LYWELSXLTX XIXELGPYXL DRXSLYVNGF XXXXXXXXTS TPGTSXVXLX
7351  TSGTPXXXPX XTXXXPLLXP FTLNFTITNL XYEEXMXXPG SRKFNTTERV
7401  LQGLLXPXFK XTSVGXLYSG CRLTLLRXEK XXAATXVDXX CXXXXDPXXP
7451  GLDREXLYWE LSXLTXXIXE LGPYXLDRXS LYVNGFHPRS SVPTTSTPGT
7501  STVHLATSGT PSSLPGHTAP VPLLIPFTLN FTITNLHYEE NMQHPGSRKF
7551  NTTERVLQGL LGPMFKNTSV GLLYSGCRLT LLRPEKNGAA TGMDAICSHR
7601  LDPKSPGLDR EXLYWELSXL TXXIXELGPY XLDRXSLYVN GFXXXXXXXX
7651  TSTPGTSXVX LXTSGTPXXX PXXTXXXPLL XPFTLNFTIT NLXYEEXMXX
7701  PGSRKFNTTE RVLQGLLXPX FKXTSVGXLY SGCRLTLLRX EKXXAATXVD
7751  XXCXXXXDPX XPGLDREXLY WELSXLTXXI XELGPYXLDR XSLYVNGFTH
7801  QNSVPTTSTP GTSTVYWATT GTPSSFPGHT EPGPLLIPFT FNFTITNLHY
7851  EENMQHPGSR KFNTTERVLQ GLLTPLFKNT SVGPLYSGCR LTLLRPEKQE
7901  AATGVDTICT HRVDPIGPGL DREXLYWELS XLTXXIXELG PYXLDRXSLY
7951  VNGFXXXXXX XXTSTPGTSX VXLXTSGTPX XXPXXTXXXP LLXPFTLNFT
8001  ITNLXYEEXM XXPGSRKFNT TERVLQGLLX PXFKXTSVGX LYSGCRLTLL
8051  RXEKXXAATX VDXXCXXXXD PXXPGLDREX LYWELSXLTX XIXELGPYXL
8101  DRXSLYVNGF THRSSVPTTS SPGTSTVHLA TSGTPSSLPG HTAPVPLLIP
8151  FTLNFTITNL HYEENMQHPG SRKFNTTERV LQGLLKPLFK STSVGPLYSG
8201  CRLTLLRPEK HGAATGVDAI CTLRLDPTGP GLDREXLYWE LSXLTXXIXE
8251  LGPYXLDRXS LYVNGFXXXX XXXXTSTPGT SXVXLXTSGT PXXXPXXTXX
8301  XPLLXPFTLN FTITNLXYEE XMXXPGSRKF NTTERVLQGL LXPXFKXTSV
```

Repeat Domain

TABLE 21 - continued

CA125 Protein Sequence

(SEQ ID NO: 162)

```
8351  GXLYSGCRLT LLRXEKXXAA TXVDXXCXXX XDPXXPGLDR EXLYWELSXL
8401  TXXIXELGPY XLDRXSLYVN GFTHRTSVPT TSTPGTSTVH LATSGTPSSL
8451  PGHTAPVPLL IPFTLNFTIT NLQYEEDMHR PGSRKFNTTE RVLQGLLSPI
8501  FKNSSVGPLY SGCRLTSLRP EKDGAATGMD AVCLYHPNPK RPGLDREQLY
8551  CELSQLTHNI TELGPYSLDR DSLYVNGFTH QNSVPTTSTP GTSTVYWATT
8601  GTPSSFPGHT XXXPLLXPFT LNFTITNLXY EEXMXXPGSR KFNTTERVLQ
8651  GLLXPXFKXT SVGXLYSGCR LTLLRXEKXX AATXVDXXCX XXXDPXXPGL
8701  DREXLYWELS XLTXXIXELG PYXLDRXSLY VNGFTHWSSG LTTSTPWTST
8751  VDLGTSGTPS PVPSPTTAGP LLVPFTLNFT ITNLQYEEDM HRPGSRKFNA
8801  TERVLQGLLS PIFKNTSVGP LYSGCRLTLL RPEKQEAATG VDTICTHRVD
8851  PIGPGLDREX LYWELSXLTX XIXELGPYXL DRXSLYVNGF XXXXXXXXTS
8901  TPGTSXVXLX TSGTPXXXPX XTXXXPLLXP FTLNFTITNL XYEEXMXXPG
8951  SRKFNTTERV LQGLLXPXFK XTSVGXLYSG CRLTLLRXEK XXAATXVDXX
9001  CXXXXDPXXP GLDREXLYWE LSXLTXXIXE LGPYXLDRXS LYVNGFTHRS
9051  FGLTTSTPWT STVDLGTSGT PSPVPSPTTA GPLLVPFTLN FTITNLQYEE
9101  DMHRPGSRKF NTTERVLQGL LTPLFRNTSV SSLYSGCRLT LLRPEKDGAA
9151  TRVDAVCTHR PDPKSPGLDR EXLYWELSXL TXXIXELGPY XLDRXSLYVN
9201  GFXXXXXXXX TSTPGTSXVX LXTSGTPXXX PXXTXXXPLL XPFTLNFTIT
9251  NLXYEEXMXX PGSRKFNTTE RVLQGLLXPX FKXTSVGXLY SGCRLTLLRX
9301  EKXXAATXVD XXCXXXXDPX XPGLDREXLY WELSXLTXXI XELGPYXLDR
9351  XSLYVNGFTH WIPVPTSSTP GTSTVDLGSG TPSSLPSPTT AGPLLVPFTL
9401  NFTITNLQYG EDMGHPGSRK FNTTERVLQG LLGPIFKNTS VGPLYSGCRL
9451  TSLRSEKDGA ATGVDAICIH HLDPKSPGLD REXLYWELSX LTXXIXELGP
9501  YXLDRXSLYV NGFXXXXXXX XTSTPGTSXV XLXTSGTPXX XPXXTXXXPL
9551  LXPFTLNFTI TNLXYEEXMX XPGSRKFNTT ERVLQGLLXP XFKXTSVGXL
9601  YSGCRLTLLR XEKXXAATXV DXXCXXXXDP XXPGLDREXL YWELSXLTXX
9651  IXELGPYXLD RXSLYVNGFT HQTFAPNTST PGTSTVDLGT SGTPSSLPSP
9701  TSAGPLLVPF TLNFTITNLQ YEEDMHHPGS RKFNTTERVL QGLLGPMFKN
9751  TSVGLLYSGC RLTLLRPEKN GAATRVDAVC THRPDPKSPG LDREXLYWEL
9801  SXLTXXIXEL GPYXLDRXSL YVNGFXXXXX XXXTSTPGTS XVXLXTSGTP
9851  XXXPXXTAPV PLLIPFTLNF TITNLHYEEN MQHPGSRKFN TTERVLQGLL
9901  RPLFKSTSVG PLYSGCRLTL LRPEKHGAAT GVDAICTLRL DPTGPGLDRE
9951  RLYWELSQLT NSVTELGPYT LDRDSLYVNG FTQRSSVPTT SIPGTSAVHL
10001 ETSGTPASLP GHTAPGPLLV PFTLNFTITN LQYEVDMRHP GSRKFNTTER
10051 VLQGLLKPLF KSTSVGPLYS GCRLTLLRPE KRGAATGVDT ICTHRLDPLN
10101 PGLDREQLYW ELSKLTRGII ELGPYLLDRG SLYVNGFTHR NFVPITSTPG
10151 TSTVHLGTSE TPSSLPRPIV PGPLLVPFTL NFTITNLQYE EAMRHPGSRK
10201 FNTTERVLQG LLRPLFKNTS IGPLYSSCRL TLLRPEKDKA ATRVDAICTH
10251 HPDPQSPGLN REQLYWELSQ LTHGITELGP YTLDRDSLYV DGFTHWSPIP
10301 TTSTPGTSIV NLGTSGIPPS LPETTXXXPL LXPFTLNFTI TNLXYEEXMX
10351 XPGSRKFNTT ERVLQGLLKP LFKSTSVGPL YSGCRLTLLR PEKDGVATRV
10451 DAICTHRPDP KIPGLDRQQL YWELSQLTHS ITELGPYTLD RDSLYVNGFT
10501 QRSSVPTTST PGTFTVQPET SETPSSLPGP TATGPVLLPF TLNFTITNLQ
10551 YEEDMHRPGS RKFNTTERVL QGLLMPLFKN TSVSSLYSGC RLTLLRPEKD
10601 GAATRVDAVC THRPDPKSPG LDRERLYWKL SQLTHGITEL GPYTLDRHSL
10651 YVNGFTHQSS MTTTRTPDTS TMHLATSRTP ASLSGPTTAS PLLVLFTINF
10701 TITNLRYEEN MHHPGSRKFN TTERVLQGLL RPVFKNTSVG PLYSGCRLTL
10751 LRPKKDGAAT KVDAICTYRP DPKSPGLDRE QLYWELSQLT HSITELGPYT
10801 QDRDSLYNVG FTQRSSVPTT SVPGTPTVDL GTSGTPVSKP GPSAASPLLV
10851 LFTLNGTITN LRYEENMQHP GSRKFNTTER VLQGLLRSLF KSTSVGPLYS
10901 GCRLTLLRPE KDGTATGVDA ICTHHPDPKS PRLDREQLYW ELSQLTHNIT
```

Repeat Domain

TABLE 21 - continued

CA125 Protein Sequence

(SEQ ID NO: 162)

10951	ELGHYALDND SLFVNGFTHR SSVSTTSTPG TPTVYLGASK TPASIFGPSA
11001	ASHLLILFTL NFTITNLRYE EN**M**WPGSRKF NTTERVLQGL LRPLFKNTSV
11051	GPLYSGSRLT LLRPEKDGEA TGVDAICTHR PDPTGPGLDR EQLYLELSQL
11101	THSITELGPY TLDRDSLYVN GFTHRSSVPT TSTGVVSEEP FTLNFTINNL
11151	RYMAD**M**GQPG SLKFNITDNV MKHLLSPLFQ RSSLGARYTG CRVIALRSVK
11201	NGAETRVDLL CTYLQPLSGP GLPIKQVFHE LSQQTHGITR LGPYSLDKDS
11251	LYLNGYNEPG LDEPPTTPKP ATTFLPPLSE ATTAMGYHLK TLTLNFTISN
11301	LQYSPD**M**GKG SATFNSTEGV LQHLLRPLFQ KSSMGPFYLG CQLISLRPEK
11351	DGAATGVDTT CTYHPDPVGP GLDIQQLYWE LSQLTHGVTQ LGFYVLDRDS
11401	LFINGYAPQN LSIRGEYQIN FHIVNWNLSN PDPTSSEY
	IT LLRDIQDKVT
11451	TLYKGSQLHD TFRFCLVTNL TMDSVLVTVK ALFSSNLDPS LVEQVFLDKT
11501	LNASFHWLGS TYQLVDIHVT EMESSVYQPT SSSSTQHFYL NFTITNLPYS
11551	QDKAQPGTTN YQRNKRNIED ALNQLFRNSS IKSYFSDCQV STFRSVPNRH
11601	HTGVDSLCNF SPLARRVDRV AIYEEFLRMT RNGTQLQNFT LDRSSVLVDG
11651	YSPNRNEPLT GNSDLPF**WAV ILIGLAGLLG LITCLICGVL** VTTRRRKKEG
11701	EYNVQQQCPG YYQSHLDLED LQ

Carboxy Terminal Domain

TABLE 22

CA125 Repeat Nucleotide Sequence
(SEQ ID NO: 307)

```
   1  ACTGCTGGCC CTCTCCTGGT GCCATTCACC CTCAACTTCA CCATCACCAA
  51  CCTGCAGTAT GAGGAGGACA TGCATCGCCC TGGATCTAGG AAGTTCAACA
 101  CCACAGAGAG GGTCCTGCAG GGTCTGCTTA GTCCCATATT CAAGAACACC
 151  AGTGTTGGCC CTCTGTACTC TGGCTGCAGA CTGACCTCTC TCAGGTCTGA
 201  GAAGGATGGA GCAGCCACTG GAGTGGATGC CATCTGCATC CATCATCTTG
 251  ACCCCAAAAG CCCTGGACTC AACAGAGAGC GGCTGTACTG GGAGCTGAGC
 301  CGACTGACCA ATGGCATCAA AGAGCTGGGC CCCTACACCC TGGACAGGAA
 351  CAGTCTCTAT GTCAATGGTT TCACCCATCG GACCTCTGTG CCCACCACCA
 401  GCACTCCTGG GACCTCCACA GTGGACCTTG GAACCTCAGG GACTCCATTC
 451  TCCCTCCCAA GCCCCGCA
```

TABLE 23

CA125 Repeat Amino Acid Sequence
(SEQ ID NO: 308)

```
   1  TAGPLLVPFT LNFTITNLQY EEDMHRPGSR KFNTTERVLQ GLLSPIFKNT
  51  SVGPLYSGCR LTSLRSEKDG AATGVDAICI HHLDPKSPGL NRERLYWELS
 101  RLTNGIKELG PYTLDRNSLY VNGFTHRTSV PTTSTPGTST VDLGTSGTPF
 151  SLPSPA
```